This is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world's books discoverable online.

It has survived long enough for the copyright to expire and the book to enter the public domain. A public domain book is one that was never subject to copyright or whose legal copyright term has expired. Whether a book is in the public domain may vary country to country. Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that's often difficult to discover.

Marks, notations and other marginalia present in the original volume will appear in this file - a reminder of this book's long journey from the publisher to a library and finally to you.

Usage guidelines

Google is proud to partner with libraries to digitize public domain materials and make them widely accessible. Public domain books belong to the public and we are merely their custodians. Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying.

We also ask that you:

+ *Make non-commercial use of the files* We designed Google Book Search for use by individuals, and we request that you use these files for personal, non-commercial purposes.

+ *Refrain from automated querying* Do not send automated queries of any sort to Google's system: If you are conducting research on machine translation, optical character recognition or other areas where access to a large amount of text is helpful, please contact us. We encourage the use of public domain materials for these purposes and may be able to help.

+ *Maintain attribution* The Google "watermark" you see on each file is essential for informing people about this project and helping them find additional materials through Google Book Search. Please do not remove it.

+ *Keep it legal* Whatever your use, remember that you are responsible for ensuring that what you are doing is legal. Do not assume that just because we believe a book is in the public domain for users in the United States, that the work is also in the public domain for users in other countries. Whether a book is still in copyright varies from country to country, and we can't offer guidance on whether any specific use of any specific book is allowed. Please do not assume that a book's appearance in Google Book Search means it can be used in any manner anywhere in the world. Copyright infringement liability can be quite severe.

About Google Book Search

Google's mission is to organize the world's information and to make it universally accessible and useful. Google Book Search helps readers discover the world's books while helping authors and publishers reach new audiences. You can search through the full text of this book on the web at `http://books.google.com/`

HARVARD
COLLEGE
LIBRARY

SCIENCE CENTER LIBRARY

Digitized by Google

Digitized by Google

Digitized by Google

Digitized by Google

Digitized by Google

PUBLICATIONS OF THE

310
AM37
15

AMERICAN

STATISTICAL ASSOCIATION

VOLUME XV
NOS. 113-120

1916-1917

BOSTON:
AMERICAN STATISTICAL ASSOCIATION
1918

Digitized by Google

293
6

HARVARD UNIVERSITY LIBRARY
LAMONT FUND
PURCHASED FROM
CAMBRIDGE PUBLIC LIBRARY
10 MARCH 1948

Copyrighted, 1918, by
AMERICAN STATISTICAL ASSOCIATION

67540

Digitized by Google

67540

CONTENTS

LEADING ARTICLES

	Page
The Nature and Significance of the Changes in the Birth and Death Rates in Recent Years. *Walter F. Willcox*	1
Statistics of Imports and Exports. *Frank R. Rutter*	16
The Mortality of Our Public Men. *Irving Fisher*	35
The Agricultural Element in the Population. *Eugene Merritt*	50
The Mother and Infant Mortality. *Henry H. Hibbs, Jr.*	66
Measures of Relative Tax Burdens. *A. E. James*	80
Statistics of the Food Supply in Germany. *Robert Morse Woodbury*	93
Proceedings of the Seventy-seventh Annual Meeting of the American Statistical Association, Washington, D. C., December 28–30, 1915	110
Some Problems of Population Growth. *E. Dana Durand*	129
The Budgets of Smith College Girls. *F. Stuart Chapin*	149
A Comparison of the Relative Healthfulness of Certain Cities in the United States Based upon the Study of their Vital Statistics. *G. E. Harmon*	157
The Improvement of Statistics of Cause of Death Through Supplementary Inquiries to Physicians. *Louis I. Dublin and Edwin W. Kopf*	175
Statistical Tabulation and Practice. *Roland P. Falkner*	192
Statistical Standardization in Washington. *Julius H. Parmelee*	201
Classification and Limitations of Statistical Graphics. *E. A. Goldenweiser*	205

Digitized by Google

Page

The Use of the Correspondence Method in Original Research. *Asher Hobson* 210

Scope and Method of Statistics. (*Introduction by Walter F. Willcox.*) *Harald Westergaard* . . . 225

Comment on Westergaard's "Scope and Method of Statistics." *Warren M. Persons* 277

Comment on Westergaard's "Scope and Method of Statistics." *Julius H. Parmelee* 279

Comment on Westergaard's "Scope and Method of Statistics." *Carl J. West* 282

Comment on Westergaard's "Scope and Method of Statistics." *G. P. Watkins* 288

The Federal Farm Loan Act. *Walter B. Palmer* . . 292

Wages in American Cities. *W. F. Ogburn and Esther Kelly* 313

Highway Accidents in New York City During 1915. *Mills E. Case* 318

Regularity of Employment—Need of Standardization of Methods for Measuring Regularity of Employment. *N. I. Stone* 324

A Study in Social Dynamics: A Statistical Determination of the Rate of Natural Increase, and of the Factors Accounting for the Increase of Population in the United States. *John M. Gillette* 345

Forty-five Years of U. S. Naval Morbidity Statistics of Degenerative Disorders. *Charles N. Fiske* . . . 381

Florence Nightingale as Statistician. *Edwin W. Kopf* 388

Consumption of Gas as Affected by Population. *G. P. Watkins and L. H. Lubarsky* 405

Interpolation of Ordinates and Areas among Areas. *C. H. Forsyth* 418

Charters for American Business Corporations in the Eighteenth Century. *Joseph S. Davis* 426

Digitized by Google

Page

The Application of Statistics to Advertising and Marketing. *Hyman L. Roth* 436

Do the Statistics of the Concentration of Wealth in the United States Mean What They are Commonly Assumed to Mean? *Allyn A. Young* 471

Desirable Additions to Statistical Data on Wealth and Income. *Willford I. King*. 485

The Classification of Mental Diseases in New York State. *Horatio M. Pollock* 502

The Increasing Mortality After Age Forty-five—Some Causes and Explanations. *Louis I. Dublin* . . 511

On the Correlation Between the Vote for Suffrage and the Vote on the Liquor Question. A Preliminary Study. *C. E. Gehlke* 524

The Present Status of Birth Registration in American Cities and Its Relation to the Infant Mortality Rate. *Louis I. Dublin* 533

Proceedings of the Seventy-eighth Annual Meeting of the American Statistical Association, Columbus, Ohio, December 27–29, 1916 548

The Utilization of Statistics in Business. *Julius H. Parmelee* 565

The "Ratio" Chart for Plotting Statistics. *Irving Fisher*. 577

On the Variate Difference Correlation Method and Curve-Fitting. *Warren M. Persons* 602

Social-Economic Groups of the United States. *Alba M. Edwards* 643

The Horizontal Zero in Frequency Diagrams. *Earle Clark* 662

The Coefficient of Correlation. *William Gardner Reed* 670

An Interpolation Formula for "Equidistant" Frequency Distributions. *Harry Langman* 685

Digitized by Google

Page

On the Alleged Increase of Cancer. *Walter F. Willcox* . 701

Special Meeting of the American Statistical Association 783

Should We Have an Annual Census of Agriculture and Manufacturers? *O. P. Austin* 786

Income Statistics. *M. C. Rorty* 794

Population Growth in Southern New England, 1810–1860. *Percy Wells Bidwell* 813

Wholesale Prices in the United States, 1791–1801. *Harold V. Roelse* 840

The Correlation of Historical Economic Variables and the Misuse of Coefficients in This Connection. *Willford I. King*. 847

The Correlation between a Component, and between the Sum of Two or More Components, and the Sum of the Remaining Components of a Variable. *J. Arthur Harris* 854

Persistency of Dependency—A Study in Social Causation. *Maurice B. Hexter* 860

REVIEWS AND NOTES.

American Public Health Association, Preliminary Program . . . 342

Anthropometric Statistics of Children—Ages Six to Forty-eight Months. *Frederick S. Crum* 332

A Plan for Gathering Statistical Data as a By-product of Administrative Work. *H. H. Hibbs, Jr.* 336

Appendix A, Vol. I. Census of the Commonwealth of Australia. *Wm. B. Bailey* 802

A Simple Method for Finding the Moving Average. *Jens P. Jensen* 121

Association, American Public Health, Preliminary Program . . . 342

Australia, Census of the Commonwealth of, Appendix A, Vol. I. *Wm. B. Bailey* 802

Births in South Dakota, Some Statistics Relating to. *T. E. M.* . . 558

Boards of Trade and Business Statistics. *E. W. Kopf*. 467

Calculation of Partial Coefficients of Correlation and Regression Equations, Tables to Facilitate the. *Harry Langman* . . . 695

Casual Actuarial and Statistical Society of America. *E. W. Kopf* . 219

Digitized by Google

Page

Causes of Death, Nomenclature and Classification of. *E. W. Kopf* 562

1920 Census in Its Relation to Vital Statistics. *E. W. Kopf* . . 340

Classification of Industries, Standards in Statistics of Industrial Accidents and an Advance Step in the. *E. W, Kopf* . . . 563

Coefficient of Correlation, Reed on the. *J. Arthur Harris* . . . 803

Comment on Professor Irving Fisher's Article on the "Ratio Chart." *G. P. Watkins* 805

Committees 224

Committees. *Carroll W. Doten* 699

Economics of Retailing. *Carroll W. Doten* 222

England, Medical Statistics in. *E. W. Kopf* 220

Federal Registration Service, History of the. *E. W. Kopf* . . . 562

Food Inspection Service in Philadelphia. *Carol Aronovici* . . . 466

Frequency Curves and Statistical Methods, The Mathematical Theory of Probabilities and Its Application to. *Edwin Bidwell Wilson* 468

History of the Federal Registration Service. *E. W. Kopf* . . . 562

"The Homicide Record of American Cities for 1915," The Spectator, December 21, 1916. *Raymond D. Fosdick* 559

Improvement in Trade Statistics. *E. W. Kopf* 124

Income Tax Procedure. *Carroll W. Doten* 563

Indexing and Filing. *C. W. Doten* 126

Infant Mortality, Some Recent French Statistics. *Theodore C. Merrill* 809

India. Moral and Material Progress and Condition of India During —the Year 1913–1914. *Joseph Raywid* 338

Interest Tables for Small Loans. *Wm. B. B.* 698

The Mathematical Theory of Probabilities and Its Application to Frequency Curves and Statistical Methods. *Edwin Bidwell Wilson* 468

Medical Statistics in England. *E. W. Kopf* 220

The Memorial Volume. *John Koren* 469

Minutes of the First Meeting of Philadelphia Statisticians. *Charles Scott Miller* 868

The Mortality from Cancer Throughout the World. *C.-E. A. Winslow* 330

Moving Average, A Simple Method of Finding the. *Jens P. Jensen* 121

New Statistical Books Added to Boston Public Library During Quarter Ending March 31, 1916. *Library Committee* . . . 222

New Statistical Books Added to Boston Public Library. *Library Committee* 808

Digitized by Google

Page
New York City, Statistics of Religious Faiths in. *E. W. Kopf* . . 562
Nomenclature and Classification of Causes of Death. *E. W. Kopf* . 562
Note on the Moving Average. *Percy C. H. Papps* 122
Note. *F. L. H.* 127
Orange Memorial Hospital, Forty-second Annual Report. *W. B. B.* 556
Philadelphia, Food Inspection Service in. *Carol Aronovici* . . . 466
Philadelphia Statisticians, Minutes of the First Meeting. *Charles Scott Miller* 868
Poverty and Social Progress. *C. E. Gehlke* 693
Race Problems, Select Discussions of. A Collection of Papers of Especial Use in the Study of Negro American Problems. *H. P. Fairchild* 559
"Ratio Chart," Comments on Professor Irving Fisher's Article on the. *G. P. Watkins* 805
Reed on the Coefficient of Correction. *J. Arthur Harris* 803
Resolution. *Carroll W. Doten* 223
Retailing, Economics of. *Carroll W. Doten* 222
Select Discussions of Race Problems. A Collection of Papers of Especial Use in the Study of Negro American Problems. *H. P. Fairchild* 557
Small Loans, Interest Tables for. *Wm. B. B.* 698
Social Diagnosis. *E. W. Kopf* 801
Some Recent French Statistics on Infant Mortality. *Theodore C. Merrill* 809
Some Statistics Relating to Births in South Dakota. *T. E. M.* . . 558
Standards in Statistics of Industrial Accidents and an Advance Step in the Classification of Industries. *E. W. Kopf* 563
State Censuses. *E. W. Kopf* 697
Statistical Books Added to the Boston Public Library. *Library Committee* 869
Statistical Books Added to the Boston Public Library During Quarter Ending December 31, 1915. *Library Committee* 125
Statistical Methods, The Mathematical Theory of Probabilities and Its Application to Frequency Curves and. *Edwin Bidwell Wilson* 468
Statistical Data as a By-product of Administrative Work, A Plan for Gathering. *H. H. Hibbs, Jr.* 336
Statistics of Religious Faiths in New York City. *E. W. Kopf* . . 562
Statistics of Children, Anthropometric—Ages Six to Forty-eight Months. *Frederick S. Crum* 332
Statistics, Boards of Trade and Business. *E. W. Kopf* 467
Statistics, Improvement in Trade. *E. W. Kopf* 124

Digitized by Google

Page

Storage, Handling and Sale of Food in Philadelphia. *Carol Aronovici* 466

Suicide Record for 1914, by Frederick L. Hoffmann, The Spectator, November 25, 1915. *E. W. Kopf* 123

Sweden—Historical and Statistical Handbook by Order of the Swedish Government. *Wm. B. Bailey* 120

Switzerland-Zeitschrift fur Schweizerische Statistik. *E. W. Kopf* . 339

Tables to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations. *Harry Langman* 695

The forty-second Annual Report of the Orange Memorial Hospital for the Year 1915. *W. B. B.* 556

The Social Survey. *Wm. B. Bailey* 556

Unmarried Girls with Sex Experience. *Wm. B. B.* 221

Vital Statistics, 1920 Census in Its Relations to. *E. W. Kopf* . . 340

Vital Statistics. *C. W. Doten* 871

Digitized by Google

Digitized by Google

Digitized by Google

Digitized by Google

AMERICAN STATISTICAL ASSOCIATION

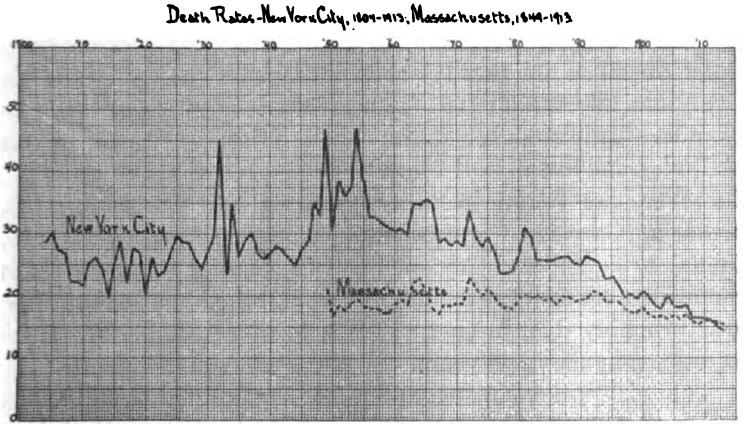
NEW SERIES, No. 113. MARCH, 1916.

THE NATURE AND SIGNIFICANCE OF THE CHANGES IN THE BIRTH AND DEATH RATES IN RECENT YEARS.*

BY WALTER F. WILLCOX, LL.D., *Professor of Economics and Statistics, Cornell University.*

Certain important changes in death rates are revealed by a diagram showing annual death rates for a populous area through a long series of years. The American districts best satisfying these two conditions are Massachusetts, for which the death rate since 1849 is known, and the old New York City, for which the death rates run back now through more than a century.

The accompanying diagram shows the death rate in each of these areas for each year from the beginning of registration to the present date.



* Paper presented at the joint meeting of Section VIII of the Second Pan-American Scientific Congress and the American Statistical Association, Washington, D. C., December 30, 1915.

Digitized by Google

This diagram illustrates what a larger body of evidence would confirm, that the perceptible and steady decline in the death rate is a relatively recent phenomenon. In order to test the correctness of this inference, the average death rate has been computed for each decade and the ratio found between that rate and the rate in the preceding ten years taken as 100. The results are as follows:

Date.	Death Rate in		Ratio of Rate to that in Preceding Decade=100.	
	New York City.	Massachusetts.	New York City.	Massachusetts.
1804–09	26.3	....	....	....
1810–19	24.7	....	94.2	....
1820–29	25.8	....	104.2	....
1830–39	29.8	....	115.7	....
1840–49	30.3	....	101.5	....
1850–59	35.6	18.0	117.7	....
1860–69	31.7	19.4	89.0	107.8
1870–79	27.6	19.7	87.2	101.6
1880–89	26.8	18.6	97.1	99.6
1890–99	23.1	19.1	86.0	97.1
1900–09	19.0	16.0	82.3	84.0
1910–13	15.3	15.5	80.6	96.7

These figures show that the decennial death rate in New York City rose for the forty years following 1810–19, but has been falling since the Civil War and is now much lower than ever before. In Massachusetts the rate rose until about 1880 and has been falling since that date, but at a slower rate than in New York. Decennial rates for at least two decades have been obtained in twenty-seven European countries and in every one the rate for 1901–10 was lower than for any previous decade. This evidence shows that the decrease of the rate is well-nigh universal in Europe. A few exceptions to the rule, however, are found in other parts of the world; namely, Ontario, Canada, Connecticut, Michigan, Vermont, Chile, Jamaica, Japan, and Ceylon. But in some of these the rates at earlier decades were incredibly low and the apparent increase was probably due, in part at least, to more complete returns of deaths in the later years.

The diagram also suggests that the annual fluctuations began to decrease at about the same time that the rates began to fall. To judge whether this is a general or only a local

Digitized by Google

change, conditions in the twenty-three countries of Europe* for which the rates as far back as 1870 are known have been examined. The average annual variation has been determined. Naturally the successive years between which the greatest variation occurred were the years of transition from peace to war or *vice versa*. Thus, the greatest changes occurred between 1871 and 1872 and between 1870 and 1871. The war of 1876 exerted a less noteworthy influence. On the whole, the annual fluctuations have decreased but the change is less obvious and uninterrupted than students of individual cities or countries might expect.

Both the steady decrease in the death rate and the slow reduction in the annual fluctuations mark and measure man's progressive emancipation from physical ills of disease or famine, which are closely dependent upon climate and season, and we might have thought, before August, 1914, from the political ills of war and revolution.

In New York State between 1894 and 1909 there was a marked decrease in mortality during July and August as compared with the average for the entire year. I believe no similar studies have been made for other states. In default of other evidence we may conclude that one factor in the diminishing death rate has been an increased control of deaths in summer. If this should prove to be a general fact, it would probably be connected with the success of efforts to reduce infant mortality. The hot months are especially dangerous to infants, while the cold months are especially dangerous to the aged.

Age influences the death rate more powerfully than any other physiological factor. Most people know that the death rate is lowest at the age of puberty and very much higher at the beginning and end of life. But the amount of difference is not commonly realized. In New York State boy babies under one year of age suffer from a death rate 65 times that of boys of 10 to 14. The death rate of nonagenarians rises to 159 times that of the youth.

It would be a natural expectation of one who is told that the death rate is rapidly falling that the gains had been distributed

* For this purpose the three divisions of Great Britain and Ireland and the eight divisions of the German Empire have been treated as countries.

Digitized by Google

somewhat evenly up and down the scale of years. But this anticipation is not borne out by the facts. Even in a hurried survey like this, two points, the possible increase of infant mortality and the possible increase of mortality among the aged, call for mention.

High medical authority in England alleged nearly ten years ago that infant mortality is stationary or increasing and this conclusion won some acceptance in the United States. Fortunately the figures proved to be susceptible of another explanation. But into that question there is now no need to go, since even in England and on the face of the figures infant mortality has decreased since 1900. Indeed among the nineteen countries of Europe for which statistics are at hand there is not one in which infant mortality has shown no decrease since the beginning of the twentieth century. In the United States the registration of births is still very defective and there are no large population groups for which we know the true infant mortality, that is, the number of children dying in the first year of life to each thousand living births. Our best substitute for this figure is the ratio between the deaths under one year of age and the living children under one. In the registration states of 1900, comprising 26.3 per cent. of the country's population, this ratio fell from 162.4 per 1,000 in 1900 to 141.7 in 1910, indicating that infant mortality fell about one eighth in the decade.

When we turn to the other end of life, the indications of our American figures are less gratifying. In the registration states of 1900 the death rate decreased between 1900 and 1910 for every age period below 55 and increased at nearly every age period above 55. On this point our experience is apparently at variance with that of Europe. The latter indicates that before 1900 the fall in the death rate as a rule extended to all ages and was especially great at ages between 5 and 53, but that for ages above 55 it was slight or in some countries absent.* The English life tables published in 1915 and speaking for a more recent period show that at every age above 5 years the mortality in 1901–10 was less than that in 1891–

* March, Statistique intern. du Mouvement de la Population, Vol. I, pp. 431–439 and 450.

Digitized by Google

1900 and that the mortality in 1910–12 was less than that in 1901–10.*

The apparent increase of mortality at high ages in our registration states, in opposition to the general trend in other countries, invites and should be given more careful and thorough analysis than it has thus far received or than I have been able to give it for this paper. Although it appears in both sexes, it does not show itself among women until the age of 60 is reached, while among men over 45 the death rate in 1910 exceeded that in 1900. It appears also among the three classes of native white, foreign born white, and colored, earliest among the colored with whom the increase appears in each sex at every age above 30, latest among the foreign born whites with whom the increase does not appear in either sex until the age of 60 and then for females appears only for the ten year age period 60–69. It appears in a prevailingly agricultural state like Vermont at an earlier age and more definitely than it does in a prevailingly industrial state like Massachusetts or Rhode Island. In a recent reference to this change, based mainly upon the figures for Massachusetts and New Jersey between 1880 and 1910, three possible causes were mentioned, "the amalgamation of the various races that constitute our population," "lack of adaptation to our rapidly developing civilization," and "some unknown biologic influence," and a preference for the second was indicated.† Thus far no evidence has been produced, I believe, showing that the increase is greatest where amalgamation of races is probably affecting the largest proportion of the population or where the lack of adaptation to a developing civilization is greatest.

An influence upon mortality which the United States is in a more favorable position to investigate is that of *race*. Under this term I do not include those residents in the United States who or whose progenitors were born in a certain country or spoke a certain language, like the Scandinavians or the French Canadians. At some future time we may be able to investigate the death rate of groups like those, although the present position and sluggish development of American vital statistics

* Reg. Gen., Supplement to 75th Ann. Rep., Pt. I, Life Tables, p. 20.

† Fisher and Fisk, How to Live, p. 282. See also Lewinski-Corwin in New York Times, Magazine Supplement, Dec. 19, 1915.

Digitized by Google

make one fear that the facts may become inaccessible before the country is ready to study them. I refer here to the two great races of white and Negro, which include between them more than 99 per cent. of our population. The registration states of 1900 contained in that year 19,544,821 whites and 388,198 Negroes and many more in 1910, groups large enough and residing in states widely enough scattered to make the changes in their death rates between 1900 and 1910 somewhat representative of the changes in the whole country.

The death rate of whites in these states fell from 17.0 in 1900 to 15.5 in 1910, a decrease of 1.5 per 1,000, or 8.8 per cent. of the initial rate. Meantime the death rate of Negroes in the same states fell from 25.6 in 1900 to 25.0 in 1910, a decrease of 0.6 per 1,000, or 2.3 per cent. These figures show that the fall in the death rate during the decade was more than twice as great among whites as among Negroes. Stating the same change in another way, the death rate among Negroes in these states exceeded that among whites by 51 per cent. in 1900 and 61 per cent. in 1910.

But to compare the two races in this way and stop without noting whether significant differences exist between the age and sex composition of the two groups might easily lead us into error. Thus among the whites 50.5 per cent., but among the Negroes only 48.3 per cent. were male. As the female death rate is regularly lower than the male, the true difference between the death rate of the races would probably be greater than the foregoing figures indicate. Even more important is the fact that of the whites 21.5 per cent., but of the Negroes only 16.1 per cent. are either children under 5 or aged 55+ and have the high death rate characteristic of infancy and old age.

Probably the best way to measure the effect of these differences in sex and age composition is to compute standardized death rates for each race. When the death rates of each race for a given sex and age are applied to the population of the same sex and age in a standard million distributed as in the registration states of 1900, the standardized death rate in 1910 is found to be 15.5 for whites and 27.5 for Negroes, showing that the corrected death rate of Negroes exceeds that of whites

Digitized by Google

by 12.0 per 1,000, or 77 per cent. The corresponding standardized death rates in 1900 were 16.9 for whites and 29.0 for Negroes, showing a fall during the decade of 1.4 per 1,000 for whites and 1.5 per 1,000 for Negroes. But in 1900 the standardized Negro death rate exceeded that of the whites by 72 per cent. as compared with 77 per cent. in 1910. The standardized death rate among female Negroes exceeds that among female whites by 61 per cent. in 1900 and by 72 per cent. in 1910, while the difference among males was 69 per cent. in 1900 and 84 per cent. in 1910, showing that the difference between the males of the two races is greater and increasing more rapidly than among the females. The death rate of each race is falling, but the gap separating the two races has seemingly grown wider. Before the Civil War, according to what evidence we have, the death rate of Negroes exceeded that of whites by 29.8 per cent. of the lower rate, and in 1890 the difference was 56.5 per cent. These earlier rates are not strictly comparable with each other or with those for 1900 and 1910, partly because they speak for different areas, neither of which agrees with the registration states of 1900, and partly because in them no attempt has been made to allow for differences in the age and sex composition of the two races or for changes in these respects. The general conclusions, however, that the death rate of each race has fallen by approximately the same amount and that the ratio by which the Negro rate exceeds the white has been rising, seem to be supported by evidence enough to make them deserve acceptance.

Still another influence upon mortality which has been recently investigated is that of *marital condition*. European figures indicated long ago that the mortality of husbands is much less than that of bachelors or widowers of the same age and the conclusion is now corroborated by figures for New York State. No doubt this difference is largely due to the selective process by which the men who marry are, on the average, at the time of marriage more healthy and vigorous than men of the same age who do not marry. Evidence that direct benefits to health accrue from married life is found in the death rate of widowers, which is much higher than that of husbands, in the death rate of Catholic celibate clergy, which

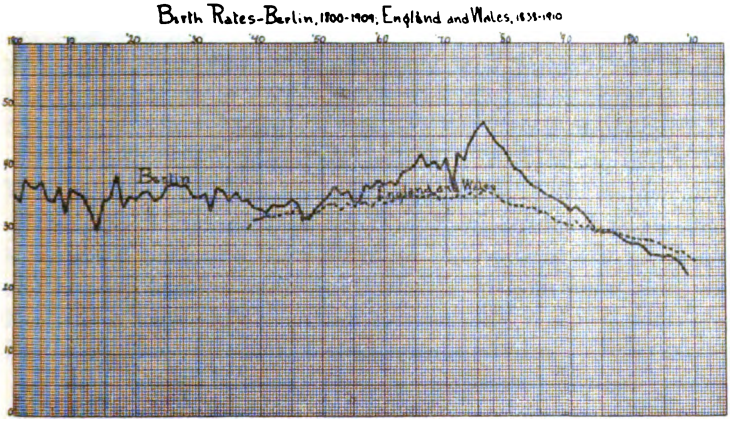
Digitized by Google

is higher than that of Protestant clergy, most of whom are married, and in the death rate of all males during the years at which marriage is most common. The normal and usual course of mortality in each sex is for a slow but steady increase to begin just after the minimum is reached in the early teens, and to continue with accelerating rapidity to the end of life. To this course there is no significant exception among women, but among men in several countries the increase is checked or even turned into a slight decrease for a longer or shorter period between the ages of 20 and 35 when marriage is most common. The probable explanation is that the normal increase of mortality during these years is nearly or quite neutralized by the steady transfer of many men from the less healthy bachelor state to the more healthy married state. The explanation finds some support in the fact that during these years the death rate of bachelors treated separately rises steadily with age. On the whole, the evidence justifies the inference that marriage exerts a directly beneficial influence on men's health and prospects of longevity. For wives the facts are less favorable. In New York the death rate of wives between 20 and 29 is greater than that of single women at the same age period. The difference is slight and part of it no doubt is due to the fact that wives 20 to 29 years of age are, on the average, more than two years older than spinsters belonging to the same age group. But as less than half the difference can be thus explained, it seems clear that in New York, as in most other areas in which the facts have been determined, during the early years of married life wives have a higher death rate than spinsters and probable that the difference is largely due to the dangers attending childbirth, especially among primiparae. But though the mortality of wives 20–29 years old is greater than that of spinsters or husbands it is less than that of bachelors.

There are no American *birth rates* which extend over a long series of years and may be trusted as accurate. In default of evidence from this country I have used in the diagram the birth rates for England and Wales since 1838 and for Berlin since the beginning of the last century. The diagram shows that the birth rate was probably at its maximum about 1875

Digitized by Google

and has decreased almost steadily since that date, that the decrease in England has been about one third and in Berlin about one half of the maximum amount, that the variations both from year to year and through longer periods were greater in the city than in the entire state, that the decrease since 1876 has been almost uninterrupted, that there was a marked depression in both areas in 1890, and that Berlin had a much more notable depression in 1871. If in the other diagram we had used the death rates for Berlin, high points on that curve would have been revealed in 1871 and 1890 when the birth rates were low.



As a rule influences which tend to increase deaths tend also to decrease births, and influences which tend to decrease deaths tend to increase births. This appears even in the rhythm of each day, Italian figures apparently showing that deaths are most frequent and births least frequent in the afternoon. There is also a yearly as well as a daily rhythm traceable in the figures but in this case the reciprocal relationship is between conceptions as mirrored in the births nine months later and deaths. There is some evidence that during the late spring and early summer and again during the late fall the death rate is low and the conception rate high. There is some evidence, likewise, that during the late winter and the late summer the death rate is high and the conception rate low.

Digitized by Google

This reciprocal relationship between births or conceptions and deaths appears also in the case of any great social calamity. As a war or a pestilence raises the death rate, so likewise it depresses the birth rate, and in estimating the social effect of either it is of the first importance to consider not only the deaths it has caused but also the births it has prevented. For example, in Massachusetts between 1860 and 1864 the death rate rose from 18.7 to 22.8, an increase of 4.1 per 1,000, and the birth rate fell from 29.3 to 24.2, a decrease of 5.1 per 1,000. In Sweden continuous and trustworthy records of births and deaths have been maintained without a break since 1749, a longer period than in any other country. The year 1773 was marked by the heaviest death rate of this century and two thirds and during that year the birth rate was also lower than in any year before or since. The most universal and fatal epidemic which has afflicted civilized countries of recent years was probably the first of the recent visitations of influenza, which spread over Europe and America in the winter of 1889–90 and caused in New York State about 5,000 deaths. Wherever the births were reported and published by months one finds, nine months after the influenza epidemic was at its height, a marked shortage of births. There were at least 200,000 fewer births in Europe in 1890 than the average annual number for the preceding five year period.

Now the most marked change in the birth rate during the last half century, a change revealed by a comparison of the two diagrams, has been the gradual decline and almost complete disappearance of this reciprocal relationship between births and deaths and the appearance in its place of a tendency for births and deaths to change in the same way rather than in opposite ways. Before proceeding to consider the causes, let me set forth the facts a little more fully.

1. The birth rate and death rate now remain approximately the same in any given country during any few years. The sharp annual variations which characterized these rates and which are still traceable in the statistics of undeveloped countries are disappearing.

2. The tendency of both death rates and birth rates is to undergo large and important modifications in longer

Digitized by Google

periods of time. The sharp up or down movements in both curves connected with such causes as war, pestilence, or famine, on the one hand, or bountiful harvests and cheap food, on the other, are being succeeded by a steady progressive downward movement in the death rate and the birth rate.

Regarding the birth rate in the United States we know practically nothing. But in default of this information I have found an available substitute by comparing the number of children under 5 years of age at the date of each census with the number of women 16–44 years of age at the same census. The results are given in the following table, in which the figures before 1850 are estimated from such data regarding sex and age as the earlier censuses afford.

Date.	Number of Children Under 5 Years of Age to 1,000 Women 16–44 * Years of Age.
1800	976
1810	976
1820	928
1830	877
1840	835
1850	699
1860	714
1870	649
1880	635
1890	554
1900	541
1910	508

During the sixty years 1850–1910 the number of children to 1,000 women of child-bearing age decreased in the United States by 191, or an average of 32 in each decade. There are only about seven tenths as large a proportion of children in the United States now as there were in 1850. If we assume that the change will continue in the direction in which it has been moving ever since 1860 and at this average rate of 32 in a decade, the number of children under 5 in the country to each 1,000 women 16–44 will be as follows:

* In order to reduce the error of the estimates in the earlier decades to a minimum, ages 16–44 were chosen instead of 15–49, the more usual limits.

Digitized by Google

1920	476
1930	444
1940	412
1950	380
2000	220
2050	60
2060	28
2070	0

The figures indicate that, if changes like those which have been in progress in the United States since 1850 should continue unchecked for a century and a half, there would be no children left. Let me not be understood as predicting a continuance of the movement for any long period in the future. But often the best method of bringing home to ourselves the vast sweep and significance of the changes revealed by statistics is to project them into the future and see whither they lead. No doubt social movements do not occur along straight lines. On the contrary sharp inflections in the curves of social change are frequent. But it is one of the main duties of statistics to point out the trend of the stream along which society is moving and thus perhaps to arouse a desire for a change.

This tendency to a decline in the birth rate is in nowise confined to the United States. On the contrary the movement in most European countries has been in the same direction. In twenty-four European countries, all, except Ireland, Portugal, and Bulgaria, for which the records are at hand, the birth rate 1901–10 was lower than in the preceding decade and in nearly all of them it was lower than in any earlier decade.

In considering the causes of this great change, let me refer first to the position of Herbert Spencer. He has argued that the various organs of the body compete with each other for nourishment and growth, that the surplus not required by the individual is all that can be devoted to the continuance of the race, that no other system makes demands upon the body as heavy as those of the nervous system, that civilization and education are steadily increasing this drain and decreasing the surplus. He finds, therefore, a natural and inevitable connection of a physiological kind between an advancing civilization and a decreasing birth rate. Some students of American statistics have sought to find support for this posi-

Digitized by Google

tion in our fragmentary and elusive material. I cannot go farther with the question this morning than to express my judgment that these efforts have not been successful and that there is no conclusive evidence, statistical or otherwise, in support of Spencer's contention. While admitting the heavy and increasing demands upon the nervous system made by modern conditions, I would point out that the decreased death rate and the decrease of sickness by which it is probably attended mean an increase of human vitality and so of the surplus to be drawn upon. Whether the increased expenditure on the nervous system equals or exceeds this increased surplus no one has even tried to prove. Until that is done I believe the Spencerian theory must be deemed only a theory.

Nor can we admit, as others have argued, that the decreased birth rate in civilized countries is due either to the growing abuse of alcohol or to the spread of venereal disease. Such arguments have come mainly from special students of these social evils and such students often lose the sense of proportion and find a relief from every social ill in the one reform on which their eyes are riveted.

Walker explained the decrease in the American birth rate by the menace to the American standard of life from the influx of swarms of immigrants accustomed to cheap food and clothing and bad housing and to the effect of this menace upon the birth rate primarily of the native stock and ultimately of the entire population. This explanation is improbable, because the decrease as we have seen began as early as 1810, when immigration was an unimportant influence, and has been matched in Australia, where it must be due to other causes than that assigned by Walker, since Australia has had no great influx of immigrants.

Turning from these inadequate explanations, the true reason for the fall in the birth rate is that in modern times, mainly within the last half century, births and the birth rate have come under the control of human will and choice in a sense and to a degree never before true. Our leading American authority, Dr. John Shaw Billings, put it as follows: "The most important factor in the change is the deliberate and voluntary avoidance or prevention of child-bearing on the part of a

Digitized by Google

steadily increasing number of married people who prefer to have but few children." Before this change began the birth of a child was, to be sure, the result of normal physiological processes, but in the vast majority of cases the birth itself did not indicate a deliberate preference for that result on the part of both or either of the parents. There is not a single one among the experts who denies that this is the great underlying cause of the modern decline in the birth rate of all civilized countries.

In considering this change may I first suggest that some such change was an almost necessary consequence of the great decline in the death rate? That is, if the death rate in Europe had declined as rapidly as it has and the birth rate had not declined, the population of that continent would now be increasing even faster than the wealth or the food supply. The standard of living would be sinking and we would probably soon relapse into our former ill state. It is the decline in the birth rate, and only that, which has enabled mankind to grip and hold fast the advantages promised by the decline in the death rate.

But there is a very important difference between the two changes. It is probably to the interest of society in the long run that each individual should be given a chance to live out his life to old age, and social effort directed to that end is beneficial both to the individual and to society. Thus far the interests of the two coincide. For this reason the two have coöperated and are coöperating effectively to reduce the death rate. But in the matter of the birth rate there is a lack of adjustment between the interests of society and those of the individual. Society is deeply concerned that enough children should be born to secure its own permanence and a reasonable increase and that those children should have the highest promise of service. The individual is deeply concerned not to compromise his own future by assuming responsibility for wife or family without the prospect of being able to maintain them in accordance with his standard of living. The individual may often see for himself or herself, therefore, a balance of advantage in abstinence from or postponement of marriage, in a childless marriage or a small family, while society from its point of view might conceive it to be most important that a given endowment of much social worth should be perpetuated.

Digitized by Google

If there were time it would be easy to show that a low and diminishing birth rate is especially characteristic of many strains of population, like college graduates of both sexes and the native American stock of the New England states, stocks perhaps better endowed than the average population with hereditary qualities the perpetuation of which is socially desirable.

While persons engaged in grappling with public health problems should interest themselves in the various changes I have briefly outlined, the main question which my figures raise is this: How shall the desirable natural increase of the population be secured and at the same time the quality of the population be maintained or improved by securing at least a normal or average and, if possible, a more than normal birth rate and natural increase in the strains of population which are of the best stock and therefore likely to transmit qualities of greatest social worth?

In this difficult field a few general principles may be stated dogmatically, which I would be glad to explain and defend, if there were time.

1. The death rate cannot be expected to fall much below where it now stands in healthy districts.

2. There is no such natural limit to a fall in the birth rate.

3. The spread in the volitional control of the birth rate is a change against which, even if we believe it undesirable, it is hopeless to struggle.

4. Legal regulations of marriage in the effort to diminish the number of births of diseased or otherwise undesirable children seem likely, unless accompanied by segregation, to do more harm than good.

5. The social service rendered by parents who have hereditary qualities of great value and make heavy sacrifices in other directions in order to rear families of normal size or larger is likely in future to be much better appreciated and requited.

6. Persons interested in maintaining the numbers and improving the quality of the population should aim not merely or mainly at a continued reduction of the general death rate but also at the gradual education of public opinion towards a readjustment of the birth rate in various classes which will enable society to gain from its best strains more than it can do under present conditions.

Digitized by Google

STATISTICS OF IMPORTS AND EXPORTS.*

By Frank R. Rutter, *Assistant Chief, Bureau of Foreign and Domestic Commerce.*

The collection of statistics of imports seems a simple task. Dutiable imports must be declared and examined for purposes of revenue. Free imports must be declared and inspected, at least in a cursory way, to make sure that they are entitled to free entry. Even exports, though subject to no duty or tax, must pass through our ports or our customs districts on the land frontier, and it is obviously easier to take a count of goods passing a given point than to count them at the numerous points of production or manufacture.

The problems involved in the collection of commercial statistics are doubtless simpler in many respects than those met in the field of census statistics, labor statistics, and crop estimates. It is not necessary to go to the man, as in the enumeration of the population, or to the factory, as in the census of manufactures; nor is it necessary to make sure that the returns obtained from a few selected observers are typical of crop conditions throughout the country. Yet there is a serious problem in import and export statistics—a problem quite distinct from those in other branches of statistics and one that has not received from American statisticians the attention that it deserves.

Customs control does not, even in the case of dutiable imports, afford an adequate check of accuracy. The primary duty of customs officers to attend to questions of revenue unavoidably and unconsciously diverts their attention from statistical requirements. Commercial statistics have been described as a customs by-product. This phrase aptly brings out the subordinate part that statistics play in the customs service. A statistical check, entirely separate from the check on the duties, is needed all along the line.

It will probably conduce to clearness to point out three familiar differences between the statistical system of Great Britain

* Paper presented at the annual meeting of the American Statistical Association, Washington, D. C., December 29, 1915.

Digitized by Google

and of the leading Continental nations. There is, first of all, a distinction between "special" and "general" commerce in the statistics of Continental countries that applies equally to imports and exports. In theory an article enters "general" commerce when it crosses the frontier; in practice the meaning is somewhat contracted, since most countries exclude, even from this most extended category, the direct transit trade and the transshipment trade. "Special" commerce, on the other hand, covers only that part of the "general" commerce that the country carries on on its own account. In "special" exports only goods that are produced within the country are included, and in "special" imports only goods intended, so far as the customs records disclose the facts, for consumption within the country. "Special" commerce, therefore, measures with varying degrees of success the importance of the country as a producer and consumer of goods, while "general" commerce measures the importance of the country as a trading nation.

In Great Britain, as in this country, the distinction is made more directly, but less logically. British exports are divided into "British goods" and "foreign or colonial goods"—just as we divide exports into "domestic exports" and "foreign exports." To obtain the total of the "general" exports according to the Continental method, these two classes must be added together. On the import side, British and American practice approaches more closely the Continental; goods become "imports" as soon as they cross the frontier, but do not become "imports for consumption" until they are released from customs control and are free to circulate throughout the country.

A second point of difference is the greater emphasis placed on the distinction between free and dutiable goods in the British and American statistics than in the Continental. Doubtless many members of this Association have been surprised and annoyed to find that it is difficult, if not impossible, to ascertain from the commercial statistics of many Continental countries the amount of duty collected on a given article or group of articles. The point may seem to be of too little consequence to mention; it is important, I take it, because it

Digitized by Google

illustrates the official view of statistics still prevailing here and in Great Britain, as contrasted with the commercial view of statistics on the Continent. In our statistics every article has been marked "free" or "dutiable" throughout the tables of imports for consumption, imports, and (prior to July 1, 1915) foreign exports. Aside from the imports for consumption, this distinction seems of no real commercial importance, and in the case of foreign exports it is confusing and, at times of tariff change, actually misleading.

The third and most important point of difference between British and Continental statistics is the requirement of a declaration of value, as well as quantity, by the British customs. Values, both of imports and exports, must be declared by the importer and shipper in all cases here as well as in Great Britain, while on the Continent it is sufficient in most cases for the importer or shipper to declare the quantity, from which the value is calculated annually (or more frequently) by means of average values. There has been some tendency, at least in Germany, towards the system of declared values. In 1909 the number of items for which value as well as quantity must be declared before the customs was increased from 144 to 1,364 in the exports and from 18 to 32 in the imports. The larger use of the ad valorem duty in this country and until recently in Great Britain and the larger use of the specific duty on the Continent is undoubtedly responsible for the varying emphasis laid on value and quantity in commercial statistics.

LEGAL BASIS FOR STATISTICS.

Our statistical system is derived from the old British system. On the import side the requirements have to do primarily with the collection of revenue, and statistical returns are obtained incidentally. At first, an entry was the only formal document required, but this was soon supplemented by compelling the importer to produce a commercial invoice as evidence of the accuracy of his entry. By law of July 14, 1862, the invoice must be certified by an American consular officer.

The legal basis of our system of export statistics dates from 1820. In that year a law was passed requiring shippers to present manifests showing the quantity and value of their

Digitized by Google

goods but imposing no penalty upon them for failure to comply with the law. The master of the exporting vessel was likewise required to present a manifest for the whole cargo showing much the same data. If he failed to comply with this requirement clearance might be refused and if he sailed without clearance he was liable to a penalty of $500.

This system doubtless worked well when the ship master had personal knowledge of the cargo carried. He was then in a position to check the accuracy of the shippers' statements.

Under modern conditions this system does not work. The master of the vessel no longer knows the cargo. Neither he nor the agent of the line can verify the statements of the shippers. His statement of value merely reproduces the statements of the shippers and affords no test of accuracy. From a port like New York where most of the exports are shipped in liners, the vessel must sail on schedule time and an element of haste is introduced that renders it difficult, if not impossible, even to make sure that all shippers' manifests have been furnished.

It is therefore not strange that at our large ports vessels have been permitted to clear without presenting the papers required by law, on the assurance that a complete manifest would be furnished within four days after sailing. Shippers' manifests are frequently missing, yet the penalty is not imposed on the master of the vessel, doubtless because it is recognized that the fault is not primarily his but the shippers'.

At best the system had to do only with exportation by water. With the growth of railroads exports by land to Canada and Mexico assumed large volume, and prior to 1891 were very inadequately returned. In that year a law was passed requiring shippers' manifests and car manifests based in part on the regulations concerning exportation by water. A new law was enacted in 1893 which still regulates shipments to foreign countries by land.

While the shipper by vessel has in a great majority of cases a representative at the seaport, the exporter by rail has no agent at the land frontier. As a result the declaration had to be prepared by the interior shipper and the only practical way of transmitting the declaration to the collector of customs was

Digitized by Google

by entrusting it to the railroad company. The penalty for failure to present a manifest was not applied to the shipper but (in the amount of $50) to the agent of the railroad company who carried out of our jurisdiction goods not accompanied by a manifest.

The effect of the change in law is reflected in the exports through Buffalo to Canada. From less than $500,000 in 1890 the figures jumped to over $4,000,000 in 1895.

The next important change in law was the extension of the requirements in the case of exports by vessel to shipments to non-contiguous territory. This took place in 1902, after the effort to collect returns without due authority of law had been proved impossible.

EFFORTS TOWARD IMPROVEMENT.

For many years it has been realized that the method employed in collecting both import and export statistics needed a thorough revision. Departmental and inter-departmental committees have investigated the subject and have made illuminating reports and presented recommendations. Two committees—one of Treasury officials headed by Mr. N. I. Stone, and one of officials of the Department of Commerce and Labor, headed by Mr. E. Dana Durand—were appointed in 1911, and made a thorough investigation of the subject.

The chief complaints from business men are concerned with the lack of fullness and timeliness in statistical publications. The lack of accuracy in statistics of imports, as well as exports, is not generally realized, yet this is the fundamental defect and gain along this line must precede any elaboration of the present classification. There are five lines along which improvement is needed: (1) Accuracy; (2) timeliness; (3) classification; (4) presentation of results; (5) comparability.

ACCURACY OF IMPORTS.

Among the public there is now a false sense of security in the correctness of import and export returns, because they constitute a by-product of customs administration. Customs officials, however, in their zeal to safeguard the revenue are

Digitized by Google

prone to allow their attention to be diverted from those statistical data that do not directly affect the amount of import duty. There is considerable error in the statistics of dutiable imports; more in the statistics of free imports; and most of all in the statistics of exports.

At the same time it is believed that the present extent of error in import statistics has been exaggerated. The results of the test made by the former statistical committees covering transactions at New York for one week in August, 1910, have been described as "appalling." The two sets of figures then obtained were tabulated respectively from the "entry" and from the "liquidated returns." The first of these tabulations represented the facts disclosed by the entry as submitted by the importer or customs broker; it did not represent the statistics as corrected for obvious errors by the statistical office of the customhouse. On the other hand, the so-called "liquidated returns" represented not merely the changes made on the entry by the liquidation clerks, but involved a careful reëxamination of the invoice. The basis of the comparison, therefore, was between returns on the one hand less accurate than those actually compiled, and on the other hand returns more accurate than could be compiled from the entry alone, even after liquidation. The old test, however, served a most useful purpose. It brought out clearly the substantial inaccuracy of our import statistics, and, in particular, of statistics more detailed than the tariff calls for. And it brought to statistical clerks in the customhouses a wholesome distrust of the unverified statements of importers and exporters.

There is imperative need of a new test. The larger use of ad valorem duties under the present tariff necessitates more frequent consultation of the invoice since a statement of quantity, which no longer affects the amount of duty, is often omitted. Strong representations have been made recommending a new test at New York, but the unusual pressure of work due in part to the increase in exports, and in part to the efforts toward perfecting the system of mechanical tabulation, have up to the present prevented the test.

It is, however, already known that the chief errors in import statistics result from inadequate detail in the entry, and failure

Digitized by Google

to take account of corrections made by the appraiser and weigher. The first of these errors is due to lack of a statistical examination of the entry before it is accepted. The examination of the entry is a hurried one and is made by clerks more familiar with the tariff than with the statistical classification. It is, therefore, entirely natural that they should give first attention to factors that affect the revenue. The remedy, as pointed out by the Statistical Committees of 1911, is to provide a separate statistical examination of the entry. While plans are being made to put this reform into practice, it is being postponed, in part for administrative reasons and in part to await the full operation of the new export regulations.*

Whether or not the changes made in liquidation are of sufficient importance to be incorporated in the statistics of a following month has not yet been determined. Liquidated returns are used for statistical compilation in most of the minor ports, but not at New York, Boston, Philadelphia, Baltimore, New Orleans, or Chicago. At New York, liquidation is always from three weeks to six weeks, or even longer, in arrears; the change would therefore have to be made in a subsequent month; and in most cases the statistical changes noted on the entry by the liquidator are not material.

As an ideal, returns should be compiled from the invoice after appraisement. The invoice, however, is a cumbrous document, describing each lot, and sometimes each parcel, in minute detail, and usually with values in a foreign currency. The entry, on the other hand, is a summary of the invoice, and is a far more convenient document from which to compile statistics. For convenience and economy, therefore, it must be the document used, and it can be made a safe guide if the proper statistical examination is made before its acceptance.

ACCURACY OF EXPORTS.

Defects in the statistics of exports may be ascribed chiefly to lack of appropriate penalties to insure a declaration for each shipment. The present law is archaic, penalizing the vessel rather than the shipper who fails to fulfill the legal require-

*Since this paper was written, some entry clerks at New York have been designated to examine very entry for statistical details.

Digitized by Google

ment. It is most desirable also to have the original shipper, who best knows the facts, make the declaration.

The Committees of 1911 recommended a law compelling the original shipper, if in the interior, to present with the goods a declaration to be forwarded by the common carrier to the seaport, and there delivered to the Collector of Customs. Penalties were provided alike on the shipper who failed to present the declaration and on the carrier who accepted a shipment for export without the declaration. The purpose of the bill was excellent, but its provisions, in my judgment, were too drastic. It has long been known that for many shipments, especially from New York, no declaration has been presented, and, consequently, that the goods are omitted from our record. Last spring efforts were begun, both by the Statistical Bureau of the New York Customhouse and by the Marine Division, to notify delinquent shippers that declarations must be presented. Much improvement resulted, but this practice is merely a makeshift.

A more serious effort to accomplish as much as possible by administrative action was made in September, 1915, when the Secretary of the Treasury and the Acting Secretary of Commerce issued new export regulations (T. D. 35708). The purpose of the new regulations is to compel the presentation of a declaration, preferably from the original shipper, for every export shipment. A vessel is not to be allowed to clear until declarations from all shippers have been presented to the Collector of Customs at the port of exportation. In the absence of any declarations, the vessel may clear only if it gives bond to furnish within fifteen days all the missing declarations. The bond is in the sum of $500, and of this amount $50 is to be deducted for each declaration not presented within the fifteen days allowed. The regulations aroused a storm of protest and in some minor respects they were amended by order of the Secretary of the Treasury and the Secretary of Commerce on December 14, 1915 (T. D. 35969), and the date on which they became effective was postponed from January 1 to February 1, 1916. In the revised draft less insistence is placed upon the requirement that the original shipper make out his own declaration. If properly

Digitized by Google

authorized, an agent at the seaboard may perform this task. The essential thing is the responsibility of the shipper himself, and the presentation of a proper declaration.

I am in hopes that much will be accomplished by the new order. I am inclined to think, however, that it could be accomplished more easily and more satisfactorily by a new law imposing on the shipper moderate penalties: (a) For failure to file a declaration; (b) for incorrect declaration; and (c) for failure to show the necessary statistical details. This measure should contain a provision authorizing the Collector of Customs to call for the original invoice in all cases to determine whether or not the declaration is correct.

TIMELINESS.

Timeliness is the first essential, in the view of the average business man. The importance of this factor is, of course, greatest in the case of monthly statistics. Yet for the last year there has been greater delay in the publication of the "Monthly Summary of Foreign Commerce," while the annual reports are being printed more promptly than in the past. The worst record is that for October. The last statement required for the compilation of the October issue was not received by the Bureau of Foreign and Domestic Commerce until December 15, more than six weeks after the close of the month. The war is largely responsible for the delay in the receipt of monthly returns from the Collectors of Customs. Their statistical forces are in many cases not flexible and the number of men compiling the export returns has not been sufficient to keep abreast of the great increase in the number of export transactions.

The special effort to publish returns more promptly has, in itself, been a source of delay. The installation of mechanical tabulation seemed to offer the best means of expediting the returns without increasing the expenditure for statistical work. Since July 1, two sets of returns have been compiled in the various customhouses, one for transmission daily to the New York Customhouse for mechanical tabulation, and the other serving as the basis for the monthly returns to the Bureau of Foreign and Domestic Commerce. Obviously, the double tabulation

Digitized by Google

has caused delay, yet this was unavoidable since, as experience has proved, reliance could not be placed from the start on the results compiled on the tabulating machines.

The new system offers at least three advantages. The first, perhaps obvious, is the receipt of prompter returns from the customhouses. The gain will probably be found, when the new system is perfected, rather in the more prompt tabulation of the returns when the month is over than in the saving of time in the original compilation.

The second advantage to be derived from the use of tabulating machines is the elimination of much purely clerical labor—and the gain from this will be found chiefly in the more careful scrutiny of the returns that will be made possible. It is imperative to insist that the clerks relieved from clerical drudgery shall be so employed as to lessen the element of error in commercial statistics rather than to effect a smaller expenditure for statistical work. Accuracy is largely the reward of adequate care and labor, and much of the error that has crept into statistics in the past is the result of an insufficient statistical force in the customhouses.

The third advantage from the new system is its greater elasticity, and the consequent ease with which additional details can be compiled. The labor involved in increasing the classification when the returns are tabulated mechanically is far less than when compiled in blotters. The number of cards determines the time required for mechanical tabulation, and the number of cards does not increase proportionately to the number of items in the classification.

Up to the present, the results obtained by mechanical tabulation have been disappointing. It is not known whether this is due to inadequate equipment, to a lack of proper checks as to accuracy, or to a failure on the part of the New York officials to realize fully the data required by the Bureau of Foreign and Domestic Commerce. It is hoped that the difficulties can soon be overcome, for mechanical tabulation seems to be the most practical means of increasing the timeliness of our publications without sacrificing important details.*

* Substantial progress has been made since this paper was written, through closer coöperation between the Bureau and the customs officials.

Digitized by Google

CLASSIFICATION.

With the increasing use of commercial statistics by business men, great emphasis is placed upon the need of specific information. To be of value, returns must be given in greater detail. The business man is interested in a particular line, and finds of little or no help returns that combine his products with those of several related lines. With every increase in the classification, there is greater liability to error. The entry that describes merchandise sufficiently to determine the proper rate of duty is likely to be passed, even though the statistical sub-class is not shown. Many of the errors disclosed by the former test were due to this cause. In the interest of accuracy, therefore, it is out of the question to expand materially the import classification until provision is made for a preliminary statistical comparison of the entry with the invoice.

The greatest need for an expansion of the classification is found in the case of exports. The most detailed classification of exports now covers less than 600 items, while in the imports for consumption there are about 3,000 distinct items. The chief preventive of an increase in the number of items is the indefinite character of export declarations. So many articles are described mèrely by general terms that it is out of the question to separate articles frequently of much commercial importance.

Defects in the present classification, aside from its incompleteness, are the incomparability of the import and export schedules and the failure to conform to current commercial terms. The latter defect is due to the preservation in the tariff of many terms now obsolete, and the necessity of having the statistical classes follow closely the tariff items.

The whole subject of classification demands most careful investigation and the Bureau is planning to undertake this task in coöperation with commercial organizations throughout the country, as soon as the success of mechanical tabulation is assured. It is apparent that it would be unwise to study this question until the limits fixed under the new system can be ascertained.

Digitized by Google

PRESENTATION OF RESULTS.

The best method of presenting statistical results is particularly a subject on which the advice of this Association would be of the greatest value. Statistical reports continue too long to follow the same form after the conditions that first determined that form have passed away.

The statistical year is still the fiscal year extending from July 1 to June 30. This seems a relic from the period when commercial statistics were instituted for the information of the government. Now that they are no longer regarded as solely or even chiefly of interest to official classes, the period to which they relate should be that most convenient and least confusing to the public.

The business year differs greatly from industry to industry and even from establishment to establishment. So far as I have been able to ascertain, the calendar year, more largely than any other, is used for business accounts. This practice has been intensified since the enactment of the income tax. Census returns of manufactures are collected chiefly for the calendar year. The calendar year is the period for which commercial statistics of most foreign countries are published. The use of any other period tends to confusion since it is impossible in all references to describe the period as "year ending June 30."

It may be urged in favor of the present fiscal year as a period for commercial statistics that it coincides more nearly than the calendar year with the natural or crop year. The returns of wheat exported during the year ending June 30, 1915, for example, cover approximately the crop harvested in *1914*, while on the calendar year basis exports from that crop would be divided over two distinct years. With the increasing industrial character of the United States, however, this consideration becomes of less and less importance. The great users of commercial statistics are the manufacturers and traders rather than the farmers. It would always be desirable, I take it, to include in the introduction or as an appendix to the annual statistical report a statement by fiscal years, or better by crop years, of the trade in several of our staple crops.

Digitized by Google

The gain in clearness and in comparability with census statistics and foreign commercial statistics is sufficient in my judgment to render a change to the calendar year highly desirable. To effect this a change in law is necessary.

What is generally understood by the term "imports"? Legally, an article is imported when landed, whether for immediate consumption or for storage in bonded warehouses. From an economic point of view, however, bonded warehouses may well be regarded as foreign territory. The door of the bonded warehouse is really the economic frontier of the country.

Since the United States is not a large reëxporting country, the difference between "imports" and "imports for consumption" is largely one of time. The instances in which goods are exported from warehouses are few as compared with the instances in which after the lapse of time goods are entered for consumption within the country.

Perhaps the distinction is most clearly brought out by an illustration. While the last tariff act was under discussion wool in large quantities was landed at our ports and stored in bonded warehouses until December 1, 1914, when it could be withdrawn without payment of duty. Was such wool really imported when it was landed or when it was removed from the warehouse?

On the export side we have a clear distinction between domestic exports and foreign exports. On the import side imports for consumption are most nearly comparable with domestic exports, yet not fully comparable, since free goods are not generally warehoused and may be entered for consumption although intended for reëxportation. To be strictly accurate, dutiable imports for consumption should be compared with domestic exports and free imports with domestic and foreign exports combined.

I am in favor of adopting as the basis for the statement by countries imports for consumption instead of imports—first, because imports for consumption correspond more closely to the economic concept of imports; secondly, because they are more comparable, at least in the case of dutiable goods, with domestic exports; thirdly, to facilitate comparison with

Digitized by Google

foreign statistics; and last, to avoid the confusion resulting from two sets of returns, the distinction between which is not generally understood, while the difference in most cases is too small to have any real significance.

How far is it desirable to continue the distinction between "free" and "dutiable" imports and foreign exports? In the case of foreign exports, it can be of no interest to the reader to know that any particular article would be free or would be dutiable if it were imported. At a time of tariff change, the distinction is really confusing. If "free foreign exports" means anything to the reader, it means that the article shipped abroad was admitted free when it was imported. Yet an article subject to duty when imported under a preceding tariff would be recorded as free if reëxported after the article had been transferred to the free list.

The distinction seems of little value even in the case of imports by countries. In a country statement, the reader is interested in knowing the commercial movement of the article irrespective of its tariff treatment. The reader would find more convenient, I am sure, a statement of flax imported for the whole of the fiscal year 1914, than the two statements published in Commerce and Navigation showing the imports separately under the Payne-Aldrich and the Underwood Acts.

The detailed statement of imports for consumption shows not only whether articles are free or dutiable, but the rate and amount of duty on each item. This one table seems sufficient for showing the tariff treatment of goods. To lump together free articles and articles dutiable at 5 per cent. seems no worse than to group articles dutiable at 5 per cent. and 100 per cent.

As far, however, as the Bureau has felt justified in going, is to eliminate the distinction in foreign exports. This was done on July 1, 1915. If approved by appropriate Congressional committees, by business organizations, and by leading statisticians, the distinction will be discontinued likewise in the monthly returns and in the statement by countries in the annual report.

Is the basis of valuation for imports correct? This country stands with a minority of the leading nations in showing the

Digitized by Google

value in the chief markets of the foreign country where purchased. Freights, commissions, and other non-dutiable charges are excluded. As a result, our valuations are not comparable with those in the United Kingdom and most other European countries where both exports and imports are valued at the national boundary.

Perhaps the most striking instance of the unfortunate result of our method of valuation is seen in the import prices of rubber. Notwithstanding the improvement of plantation rubber, Para rubber is still quoted at a slightly higher price. In Brazil, however, there is a heavy export duty, which constitutes an important element in the price. This duty is not included in our statistical valuation with the result that the value of India rubber imported from Brazil during the fiscal year 1914 averaged only 40 cents a pound, while the import value of that from Ceylon averaged 60 cents a pound.

From a statistical point of view, the valuations should represent the value at the frontier—in the case of exports, f. o. b. values, in the case of imports, c. i. f. values. The change suggested, however, can not be effected without grave inconvenience so long as our ad valorem duties continue to be based on foreign values.

COMPARABILITY OF STATISTICS.*

That the statistics of foreign trade of different countries are far from being comparable is generally recognized, but few realize the extent to which such incomparability exists. Four of the causes responsible for this condition are: (1) Varying definitions of the terms "imports" and "exports"; (2) different methods of determining weight and value; (3) inaccurate declaration of origin and destination; and (4) differences in classification.

Varying Definition of "Imports" and "Exports." Attention has already been called to the distinction in most continental countries between "special commerce" and "general commerce." International comparisons are usually based on "special" imports and exports. Discrepancy is often caused

* The discussion of international statistical comparisons is based on my report on Uniform Classification of Merchandise, published by the International High Commission, Pan-American Financial Conference, 1915.

Digitized by Google

by the inclusion in some countries of imports of raw materials for reëxportation in manufactured form, and similarly by the inclusion on the export side of the articles so manufactured and reëxported, while other countries exclude both these classes.

Different Methods of Determining Weight and Value. While in general the weight shown statistically is intended to represent net weight, there are numerous exceptions to this rule. In the case of value the resultant discrepancies in commercial statistics are much more apparent. In the majority of countries the effort is made to show the value of exports when they leave the country and the value of imports when they reach the customs frontier, inclusive of freight and insurance, but exclusive of import duty. The United States is an exception since, in the case of imports, the foreign value (f. o. b. port of lading) instead of the value on arrival (c. i. f.) is prescribed by law. Freight charges are not, as a rule, sufficient to cause a marked discrepancy, but the exclusion of certain non-dutiable charges, such as export duties in the country of origin make our statistics of, say, imports of Brazilian rubber incomparable in point of value with the British statistics of imports of the same commodity.

Inaccurate Declaration of Country. The exports of raw cotton from the United States to Germany in the calendar year 1913, according to our own returns, amounted to 1,000,000 bales more than the German returns record. Russia, on the other hand, credits this country with more than twice the amount that our statistics show. The large transshipment trade through the free port of Hamburg and other German ports explains the first of these discrepancies. Shipments declared for Hamburg, without indication of any further destination, are of necessity recorded in our returns as exported to Germany, while in the German returns only the cotton entered for consumption is recorded. In the case of Russia, the reason for the discrepancy is much the same. Our commercial relations with that country are largely indirect, and as a result large quantities of American cotton for Russian consumption are shipped through other European ports, largely for account of firms located in such ports. Such ship-

Digitized by Google

ments have to be recorded in our returns as exported to the countries in which the transshipment ports are located. If in the Russian returns all the cotton imported could be traced back to its country of origin, the discrepancy would undoubtedly be much larger.

As a general rule, import statistics are more trustworthy than export statistics. In the case of exports, the ultimate destination of the goods is frequently unknown; while in the case of imports the country of origin, although sometimes difficult to trace, can usually be ascertained if sufficient effort is made.

Differences in Classification. The primary basis of classification is sometimes found in the origin of the product or its degree of manufacture, as in France; sometimes, as in most South American countries, the raw materials and their various manufactured forms are grouped together in great classes, such as "textiles"; sometimes, as in Germany, both these divisions are made; sometimes the articles are arranged in alphabetical order. The component material determines the classification of articles in some tariffs, while in others regard is had to the use of the article. Commercial statistics follow the tariff in practically all countries and follow it slavishly where commercial interests are not sufficiently organized to demand additional details. Statistical classification is so intricate and so varied that international comparisons are often impossible and always liable to error.

A most striking illustration of apparent discrepancy in statistics due to differences in classification is found in the shipments of pumps from the United States to Cuba. In the fiscal year ending June 30, 1914, the American returns recorded a value of $291,350 for "pumps and pumping machinery" destined to Cuba; while the Cuban returns show imports from the United States of pumps worth only $19,631.

A careful examination of the Cuban tariff and customs decisions brings out clearly the reason for the apparent discrepancy between the two sets of statistics. The item described as "pumps" in the Cuban statistics doubtless refers to the phrase "steam pumps" in tariff No. 218. But even those, if intended for use on farms or in sugar factories, are

Digitized by Google

admitted at greatly reduced rates of duty under the head of "Agricultural machinery and implements" (No. 216) or "Machinery and implements for sugar plantations" (No. 215). The vast majority of the pumps imported are undoubtedly declared for customs purposes under one of these two classes and appear under the corresponding item in the statistics. Pumps other than steam pumps not intended for use on farms or plantations are classified for duty purposes, and undoubtedly for statistical purposes, under tariff No. 226, "Machinery and apparatus not specified."

This instance should not be construed as reflecting in any manner on the accuracy of the Cuban returns. Without doubt numerous cases could be found in the statistics of the United States, where, owing to differences in classification, some articles would appear to be understated in comparison with the returns of other countries. The illustration from the Cuban statistics has been cited merely to emphasize the need of greater uniformity in classification, if general readers are not to be misled.

The adoption of a uniform classification would do much to make the statistics of different countries more comparable.

As far back as 1869, the International Statistical Congress advocated uniform classification. The International Customs Conference held in Paris in 1900 went carefully into the subject, and recommended that the various countries unite on a common statistical classification to cover only the most important articles, which would serve as a basis for international statistics, in addition to the ordinary statistics published by each country, according to its own classification. The International Statistical Conference of 1910 at Brussels drew up a tentative classification for this purpose, which was referred to the various governments represented for further consideration. With few amendments, this classification was adopted by the International Statistical Conference of 1913 at Brussels and incorporated in a convention, instituting the International Bureau of Commercial Statistics and providing for the annual publication by that bureau of the imports and exports of each of the contracting countries. The United States was not represented officially at these conferences.

Digitized by Google

The first step had thus been taken toward statistical uniformity when the European war broke out. It is to be hoped that upon the conclusion of peace the convention will be renewed.

CONCLUSION.

I have aimed in this paper to bring clearly before you the fact that a serious problem, worthy of your most careful consideration, is involved in commercial statistics. I have pointed out some of the differences between the British system, which we inherited, and the system on the Continent. For greater accuracy in respect to exports, more stringent regulations have recently been prescribed. Greater accuracy in imports can best be obtained by a preliminary statistical comparison of the entry with the invoice. Timeliness and fullness of detail are of great importance; means of improvement along these lines are now being worked out. In the presentation of results, I am in favor of adopting the calendar year as the statistical period, of showing by countries imports for consumption instead of general imports, and of discontinuing the distinction between free and dutiable goods except in the detailed statement of imports for consumption. To lessen the serious incomparability that exists between the commercial statistics of different countries, the first step has been marked by the establishment of the International Bureau of Commercial Statistics at Brussels, which unfortunately was prevented from starting its work by the outbreak of the war.

Digitized by Google

THE MORTALITY OF OUR PUBLIC MEN.

Being a Study of the Mortality of the Presidents, Vice Presidents, Senators and Representatives, of the United States Compared, with Insured Lives, and Others.*

By Irving Fisher, *Professor of Political Economy, Yale University. Chairman of the Hygiene Reference Board of the Life Extension Institute.*

This paper is primarily a study of the mortality of the Senators and Representatives of the United States. As preliminary to this study, however, we first record the facts as to the longevity of the Presidents and Vice Presidents of the United States. The number of cases in the last two groups is too few, of course, to enable us to make any safe statistical generalizations. The following table gives the facts individually as to our Presidents:

LONGEVITY OF PRESIDENTS.

(1) Name.	(2) Date of First Inaug.	(3) Years Served.	(4) Age of President when Inaug. (to Nearest Year).	(5) Actual Years Lived from Inaug. to Death (to Nearest Year)	(6) Expectation of Life at Inaug. (Amer. Exper. Table) (a).	(7) Ratio of Actual to Expectation. (5) ÷ (6) Per Cent.
Washington	Apr. 1789	7 5/6	57	11	16	69
J. Adams	Mar. 1797	4	61	29	13	223
Jefferson	Mar. 1801	8	58	25	15	167
Madison	Mar. 1809	8	58	27	15	180
Monroe	Mar. 1817	8	59	14	15	93
J. Q. Adams	Mar. 1825	4	58	23	15	153
Jackson	Mar. 1829	8	62	16	13	123
Van Buren	Mar. 1837	4	54	25	18	139
Harrison	Mar. 1841	1/12	68	1/12	9	1
Tyler	Apr. 1841	3 11/12	51	21	20	105
Polk	Mar. 1845	4	49	4	22	18
Taylor	Mar. 1849	1 1/3	64	1	12	8
Fillmore	July 1850	2 2/3	51	24	20	120
Pierce	Mar. 1853	4	48	16	22	73
Buchanan	Mar. 1857	4	66	11	11	100
Lincoln	Mar. 1861	4 1/12	52	4	19	21
Johnson	Apr. 1865	3 11/12	56	10	17	59
Grant	Mar. 1869	8	47	16	23	70
Hayes	Mar. 1877	4	54	16	18	89
Garfield	Mar. 1881	1/2	49	1/2	22	2
Arthur	Sept. 1881	3 1/2	51	5	20	25
Cleveland	Mar. 1885	8	48	23	22	105
B. Harrison	Mar. 1889	4	56	12	17	71
McKinley	Mar. 1897	4 1/2	54	4	18	22
Roosevelt	Sept. 1901	7 1/2	43	living	living	
Taft	Mar. 1909	4	51	living	living	
Wilson	Mar. 1913		56	living	living	

(a) The "American Experience table of mortality" was first published in 1868. It was constructed by Sheppard Homans, principally from the experience of the Mutual Life Insurance Co. of New York. It has been, for a long time, the common standard for life insurance companies' calculations and for statistical comparisons, although it represents a heavier mortality than that of present insurance experience.

* Paper presented at the annual meeting of the American Statistical Association, Washington, D. C., December 29, 1915.

Digitized by Google

The longevity of each President is reckoned from the date of inauguration—that is, from the time of his first *being* President. Thus Washington became President at 57 years of age when the "expectation" of life, according to the American Experience table, was 16 years. He actually lived, however, only 11 years instead of the 16. That is, he lived only 11/16, or 69 per cent., of his expectation. John Adams, on the other hand, lived after inauguration 29 years, which was 2.23 times his expectation of 13 years.

Greatest Longevity.	Least Longevity.
John Adams (223)	Harrison (the first) (1)
Madison (180)	Garfield (2)
Jefferson (167)	Taylor (8)
J. Q. Adams (153)	Polk (18)
Van Buren (139)	Lincoln (21)
Jackson (123)	McKinley (22)
Fillmore (120)	Arthur (25)

The long-lived Presidents all belonged to the first half of the list and the majority of the short-lived ones belonged to the second half. These include the three assassinated Presidents, Lincoln, Garfield, and McKinley. If we group these deceased Presidents under four successive periods, approximately equal, we have the following table in which the Vice Presidents are also included:

RATIO (PER CENT.) OF ACTUAL TO EXPECTED LONGEVITY BEYOND INAUGURATION.

	Of all Deceased.		Excluding Assassinated Presidents among Presidents; and Omitting Vice Presidents Who Became Presidents among Vice Presidents.	
Periods.	Presidents.	Vice Presidents.	Presidents.	Vice Presidents.
1st period (4 Pres.—5 V. P.)....	156	137	156	100
2d period (8 Pres.—7 V. P.).....	84	94	84	76
3d period (7 Pres.—7 V. P.).....	75	67	84	69
4th period (5 Pres.—5 V. P.)....	45	42	68	44
Total.....	82	84	93	70

The table shows that the longevity of our Presidents, omitting the three still living, is only 82 per cent. of that

Digitized by Google

"expected" according to the table, which, in turn, is less than that among insured lives of the present time. Even if we omit the assassinated Presidents, Lincoln, Garfield, and McKinley, the longevity of the remainder is still only 93 per cent. of that expected by the American Experience table and less than two thirds the expectation of the insured lives of today. We note also that the longevity of our Presidents seems to be diminishing with the increasing complexity of modern life in Washington. Their average longevity in the four successive periods was 156, 84, 75, and 45 per cent. respectively, or, omitting the assassinated Presidents, 156, 84, 84, and 68 per cent. We find the same apparent tendency among the Vice Presidents. It is also interesting to note that the Vice Presidents seem to live out a smaller fraction of their expectations than the Presidents, excluding the influence of assassinations. This fact (if it be a significant fact rather than an accidental incident in so small a number of lives) would suggest that the official social and convivial life at Washington is a greater life-shortener than the heavy responsibilities of office.

The five Presidents of Yale University, for substantially the same period as that of the above table, show a longevity equal to 137 per cent. of their expectations. The consistent conformity of all the results to the hypothesis that a "simple" life is a long life and that a "complex" life is a short one is at least striking.

But as already emphasized, the number of lives in all three lists (only 50 in all excluding duplications) is so small that any statistical conclusions whatever, without confirmatory evidence, would be dangerous.

We turn now from these preliminary remarks on Presidential and Vice Presidential lives to the lives of our Congressmen—Senators and Representatives. Here, fortunately, we have enough lives, about 7,500, to make the results of undoubted significance.

In this study I have been fortunate in having the assistance of the New York Life Insurance Company to which I am indebted for its courtesy in permitting the clerical work to be done in its offices. I am also greatly indebted to Mr. Arthur

Digitized by Google

Hunter, its Actuary, for his assistance and criticism at every stage of the work.

The method employed has not been that of expected longevity used above, but that of expected mortality. This makes it possible to include the influence of the several thousand lives of the Congressmen still living. The mortality of the American Experience table was used here also as a standard of comparison, or measuring rod. Although, as already stated, this early table does not agree with the mortality experienced at present, either by the insurance companies of today* or among the general public, it still serves as a convenient basis in terms of which any actual mortality may be expressed in percentages.†

The chief data for this investigation were obtained from the Biographical Congressional Directory, and the necessary facts relating to nearly 7,500 Congressmen were entered upon individual cards. As the Congressional year dates from March 4, it was found convenient to carry the observations up to March 3, 1912, at which point they were closed.‡

In order to discover the effect of Congressional life on mortality, the following table was constructed:

* Modern experience averages about 25 per cent. less than the American Experience table. The contrast is more marked for the earlier than for the older ages.

† Thus if (for a particular year of age) the actual death rate is half the death rate expected (for that particular year of age) according to the American Experience table, and if the same were true for each other year of age, we would say that the actual mortality is half of the expected. Usually, of course, the ratios of actual to expected death rates will be somewhat different for different years of age. In that case an average of the different ratios, duly weighted, is struck. To take a specific example, we find that out of 398 Congressmen (entering the 15 Congresses ending March 3, 1911 and being 45 to 49 years of age on entering), the *actual* number who died in the second year after entering was 7 and the number *expected* to die out of 398 was, according to the American Experience table, 4.94. Therefore the ratio of actual to expected deaths (in this second year) was $7 \div 4.94 = 1.41$, or 141 per cent. In the same way, the number actually dying in the tenth year out of 275 then surviving was 3 as against 5.47 expected out of 275. The ratio of actual to expected is, in this case, $3 \div 5.47$, or 55 per cent. In this same way the actual and expected deaths are taken, not only for the second and tenth years, but for each year involved, until all have passed from observation (either by death or by surviving to March 3, 1912, at which date all observations closed). The average ratio for all these years is then found by dividing the total deaths (159) by the total of the expected (137), giving $159 \div 137 = 1.16$, or 116 per cent., as the ratio of actual to expected for the group of Congressmen in the entire category considered.

‡ Except for the cases mentioned below, where the record was not complete, the investigation covered the entire history from date of entering Congress until death, or until March 3, 1912. The date of birth was required in order to determine the age on first entering Congress. For some Congressmen this information was not given in the Directory. The number of such cases was small (785) and sufficient accuracy was obtained by assuming that they entered at an average age. In other cases the Directory did not trace the history of the Congressman after his retirement from Congress. These cases were not large in number and were included in the investigation from the date of first entering Congress to the date of last retirement from Congress.

Digitized by Google

SHOWING THE MORTALITY SOON AFTER ENTERING CONGRESS AND LATER, DIVIDED IN ALL INTO FIVE PERIODS OF LIFE, NAMELY, A PERIOD OF TWO YEARS COVERING THE (FIRST) CONGRESSIONAL TERM; A PERIOD OF EIGHT YEARS; THE SECOND DECADE AFTER ENTERING; THE THIRD DECADE; AND THE REMAINDER OF LIFE.

(All Congresses, ending March 3, 1911.)

Number of Years Elapsed Since First Entering Congress.	Actual Deaths.	Expected Deaths.	Ratio of Actual to Expected Deaths. Per Cent.
1- 2	114	194.78	59
3-10	824	833.76	99
11-20	1,329	1,247.84	107
21-30	1,360	1,344.67	101
31 and over	1,506	2,028.62	74

The above table shows strikingly that Congressmen had a low mortality immediately after entering Congress, but a greatly increased mortality as the years passed by, reaching the maximum in the second decade after entering Congress, after which there was a decrease. In the first period, of two years, the mortality is only a little over half of the expected by the American Experience table and somewhat less than the expected if more modern tables had been used. After two years (or one Congressional term) it almost reaches and during substantially the next three decades exceeds, the expected, and after that time it falls back to three fourths of the expected.

These figures of themselves give no proof, of course, of what causes are at work. They, however, fit into the following hypotheses: (1) Congressmen are a select group having unusual vitality, and therefore a low mortality at the outset; (2) Congressional life with its responsibilities, heckling, excitements, worries, indoor and sedentary life, late hours, dining, wining, and other unhygienic conditions makes a severe drain on this vitality, showing itself in a gradually increasing mortality; and (3) some Congressmen contrive to avoid or to correct these unhygienic conditions and cause a partial return toward the original low mortality, after their less careful colleagues have died. At any rate the cardinal fact is clear that some cause or causes connected with entering Congress tends to kill Congressmen. They enter with more than the average vitality and after a single term of office possess less than the average vitality.

Digitized by Google

The same conclusion, namely that Congressional life sets, for some cause, a killing pace, is drawn from the following table:

MORTALITY OF CONGRESSMEN ACCORDING TO LENGTH OF SERVICE, NAMELY: (A) 6 YEARS OR LESS, (B) MORE THAN 6 YEARS IN PUBLIC SERVICE. (a)

(15 Congresses, ending March 3, 1911.)

Length of Service.	Age on Retirement from Congress.			
	All ages.	Under 50.	50–59.	60 and over.
6 years and under.....	97	112	103	74
Over 6 years.....	120	103	124	124

(a) This investigation is confined to the group of 15 Congresses ending in 1911. Public Service includes not only service in Congress, but also any subsequent service as President or Vice President, of the United States, Cabinet Officer, Governor of a State or Ambassador to a foreign country, and similar positions under the Southern Confederacy. For this study Congressmen were entered in the investigation *after* the number of years in Public Service pertaining to each had been completed. Thus the mortality comparisons for a man who entered Congress in 1880 and had twenty five years of Public Service ending 1905, began with 1905 and not with 1880.

The above table shows that, in general, those who have been in Public Service the longer time have a mortality of 120 per cent. as compared with 97 per cent. for those who have served the shorter time. This contrast is the more marked the older the men when retiring from Congress. Among those retiring at sixty years of age or older the ones who keep in Public Service the longer show a mortality of 124 per cent. as contrasted with 74 per cent. for those who have served the shorter time. Thus the killing pace of public life is more killing to the older men than to the younger.*

The following table (by years elapsed since retirement instead of by ages of Congressmen) was made to throw further light on the subject. It is confined to the last fourth of our Congresses.

*In fact, as the first column shows, for those retiring before the age of 50 the long service men show a slightly lower mortality than the short service men. This may be the effect of selection. It is probable that in general only the hardier men either can or will choose a long career of Public Service. Everyone knows that the strain of professional life often forces men to retire; some of these men undoubtedly have died soon afterwards. The existence of this selective influence emphasizes all the more the main conclusion that the longer the public service the greater the mortality. Moreover, somewhat apart from voluntary withdrawal from an extended public career on the part of those who find they cannot keep the pace, the contrast between the mortalities of the two groups would be still greater if there were any way of transferring from the shorter service group to the longer service group, those who died before the 6 years were up, but who, had they lived, would have been classified in the long term group.

Digitized by Google

FIFTEEN CONGRESSES ENDING MARCH 3, 1911, OBSERVED FROM DATE OF LAST RETIREMENT FROM CONGRESS.

(Ratio (per cent.) of Actual to Expected Deaths.)

Length of Public Service.	Number of Years from Retirement.	
	1–10.	Over 10.
6 years and under	105	89
Over 6 years	128	104

The above table shows the same contrast—higher mortality for the men who served over 6 years as compared with those who served less, *i. e.*, 128 per cent., as contrasted with 105 per cent., and also 104 per cent., as contrasted with 89 per cent. Incidentally, the table also shows that the adverse influence of public life is, as might be expected, felt mostly soon after retirement.

In the first decade after retirement the mortality is much higher than in the next decade as a comparison of the two columns shows. This may be due largely to the fact that many retire on account of ill health, while others may lose health as a consequence of defeat. Again, as already observed, those who survive the first 10 years after retirement are presumably those who have been able to avoid, or counteract the worries and bad hygiene which have killed off their colleagues.*

Thus far we have considered Congressmen as a single group. But the same conclusions apply separately to different periods in our national history, as the following table shows:

RATIO (PER CENT.) OF ACTUAL TO EXPECTED DEATHS.

Number of Years Elapsed Since First Entering Congress.	Group of Congresses Ending				All Combined.
	1821.	1851.	1881.	1911.	
1–10	94	112	82	85	91
11–20	119	116	97	102	107
21–30	115	110	92	86	101
31 and over	85	85	59		74

*The reason why the first table giving the mortality according to time elapsing since entering Congress shows a decided fall in mortality only after the 30 year mark, while this table shows a decided fall after the 10 year mark is clear, viz.: in the last table the years are reckoned from a later date than the first, *i. e.*, from time of retirement instead of from time of entrance. In the former table those in the second and even the third decade after entering Congress, include many who had not yet retired.

Digitized by Google

The first group consists of the Continental Congresses and the first 16 regular Congresses ending with March 3, 1821, and the other three groups, each of 15 Congresses, ending respectively March 3, 1851, 1881, and 1911. The individual cases were classified in the four groups according to the date of first entering Congress. It will be observed by following the eye down any one of the columns of the table that the mortality always rises from the first to the second decade after entering Congress and falls thereafter, especially from the third to the fourth decades.

The above table also affords an opportunity to compare one historical period with another, if the eye is run through it horizontally instead of vertically. The most striking results of such comparisons are the lower mortality of the last two groups of Congresses as compared with the first two. The low mortality of the third group seems especially remarkable.

For historical comparison, however, the best data are those in which the mortality is presented by the attained ages of the Congressmen as in the following table, which for the present purpose should be read vertically:

RATIO (PER CENT.) OF ACTUAL TO EXPECTED DEATHS BY ATTAINED AGES.

Congresses.	All Ages.	Under 40.	40–49.	50–59.	60 and over.
Continental Congresses and 16 regular Congresses ending March 3, 1821.....	100	82	121	143	91
15 Congresses ending March 3, 1851.....	100	114	133	135	91
15 Congresses ending March 3, 1881.....	79	110	103	97	74
15 Congresses ending March 3, 1911.....	92	89	85	98	91

In this table we note: (1) that, in general, there has been an improvement in mortality from the earliest period of our history to modern times, the first two groups showing 100 per cent. each and the last two, 79 and 92 respectively; (2) that for the oldest age group—those dying over 60 years of age—there has been, generally speaking, no improvement, the figures being 91, 91, 74, and 91 per cent.; (3) that the Congresses of the third period of our history had an exceptionally low mortality, especially for the older ages.

Digitized by Google

It seems difficult to account for the exceptionally low mortality in the group of Congresses from 1851 to 1881. Probably there are several reasons, such as: (1) the temporary withdrawal (about 1861–1867) of the Southern contingent of Congressmen because of the Civil War (the insurance companies have experienced a higher mortality in the South than in other sections of the country and the statistics given below show the same contrast for Congressmen); (2) the stricter standard of living which prevailed during war time; (3) the inclusion in Congress of a larger proportion of men of "plain living and high thinking" during the period of and following the Civil War. The last named factor may account for the special superiority of the Congresses of the third period in respect to the older ages at which time the "wear and tear" diseases chiefly enter into mortality.

As has been previously stated, the expected deaths in the tables of this paper have been calculated by the American Experience table, which table shows a much higher death rate than is experienced by American insured persons at the present time. For instance, the present ratio of actual to expected deaths among the insurance companies of this country is about 75 per cent. of the American Experience table. Policyholders, however, have passed a medical examination, which is not the case with Congressmen. It is necessary, therefore, to turn to some other records in order to determine approximately what relationship to the American Experience table the mortality among Congressmen should have. In the Transactions of the Actuarial Society, Vol. X, there appears an analysis by Mr. E. B. Morris of the mortality experienced among Yale graduates.*

In order that a more detailed comparison with the mortality experience of Yale graduates may be made, the following tables have been prepared.

They show the mortality of Congressmen and the mortality of Yale graduates as compared with the American Experience table.

* The figures of Mr. Morris are, so far as I know, the only figures for any representative class of the population extending continuously from the eighteenth century through the nineteenth.

Digitized by Google

RATIO (PER CENT.) OF ACTUAL TO EXPECTED DEATHS.

Groups.	Attained Ages.				
	All Ages.	Under 40.	40–49.	50–59.	60 and Over.
31 Congresses ending March 3, 1851	100	98	128	139	91
Yale Graduates 1792 to 1849 of the Academic Department.....	86	125	115	92	75

RATIO (PER CENT.) OF ACTUAL TO EXPECTED DEATHS.

Groups.	Attained Ages.				
	All Ages.	Under 40.	40–49.	50–59.	60 and Over.
30 Congresses ending March 3, 1911	82	101	96	98	77
Yale Graduates 1850 to 1900 of the Academic Dept., Graduates 1852 to 1901 of Sheffield Scientific School, and Graduates 1843 to 1901 of Law School.....	76	70	81	79	85

The lower mortality of the Yale graduates is clearly brought out in the foregoing comparison.*

As to historical trends we may note such for the mortality of Yale graduates by taking the historical groups worked out by Mr. Morris. These extend back of our national history. The chief groups are three, covering substantially the eighteenth century and the two halves of the nineteenth in the following divisions: (1) the graduates 1701–1791; (2) the graduates of the Academic Department 1792–1849; (3) the graduates of the Academic Department 1850–1900; the graduates of the Sheffield Scientific School 1852–1901; and the graduates of the Law School 1843–1901.

The ratios of actual to expected deaths (according to the American Experience table) for these three groups were: 94, 86, and 76 per cent., respectively. It is evident from these figures that there has been a continuous improvement in mortality, as was the case among Congressmen. We also

* The Yale graduates of the medical school have not been included in the above tables. The experience in regard to these graduates had not been subdivided in the form desired for the above comparison. Although the mortality among medical graduates was somewhat higher than among the other graduates, the relative number of the medical graduates being small, their exclusion does not sensibly affect the comparison.

Digitized by Google

note for Yale graduates that there is an improvement as between the two periods for all ages except "60 and over" for which, on the contrary, there has been an increase of mortality, namely from 75 to 85 per cent.*

The recent cessation of improvement in middle age mortality and in some cases even the increase of mortality after middle age is a fact we encounter at almost every turn. It is even observable in insured lives in spite of improvements in the rigor of medical selection, especially for the older ages,† and in the classes constituting policy holders. The great medico-actuarial investigation conducted under Mr. Arthur Hunter showed that mortality at ages 60 and over "has shown no distinct improvement." For the three successive periods 1885–92, 1893–1900, 1901–08, the figures were 91, 98, 93 (based on a special table of mortality).‡ The tendency is, in fact, less observable among Congressional lives than in other quarters, owing probably to the fact that public men are today held more strictly to account for their personal conduct than formerly, when laxity of various kinds was condoned.

I close this paper by a geographic comparison of Congressmen from the Southern, Eastern, and Western sections § of the United States.

*Morris's figures show, for the higher ages (above 55), a decided retrogression. The mortalities for ages over 45 were:

(Dates Exact for Academic Department and Approximate for Others.)

	Ages.			
	46–55.	56–65.	66–75.	76 and Older.
1825–1849	98	74	75	70
1850–1874	60	81	86	93
1875–1900	88	211		

† See Eugene L. Fisk, M.D., "Force of Adverse Selection among Entrants at the Extremes of Life." Proceedings Assoc. Life Ins. Med. Directors, October 24, 1907. Pp. 129–170.

‡ See Medico-Actuarial Mortality Investigation, Vol. I, Actuarial Society, New York, 1912.

§ In the Southern states were included Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, and Texas. The Eastern states for this purpose were considered to be Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania, Ohio, Indiana, Illinois, Michigan, Wisconsin, Delaware, Maryland, West Virginia, and the District of Columbia. The remaining states were classified as Western.

Digitized by Google

The results of this investigation based upon the group of 15 Congresses ending in 1911 are as follows:

	Ratio (per cent.) of Actual to Expected Deaths.
Southern	100
Eastern	93
Western	80
All states	92

The foregoing is in conformity with the experience of insurance companies—namely, that the Southern states have the highest mortality, and the Western states the lowest, while the Eastern states represent about the average mortality throughout the United States.

The results of this investigation may be summarized in the following statistical generalizations:

1. The mortality of Congressmen for the two years immediately after entering Congress has been low, being very much lower than that of the American Experience table and materially lower than the mortality of Yale graduates from 1792 on.

2. The mortality of Congressmen in the next eight years of life was much higher, being about equal to that of the American Experience table and above the mortality of Yale graduates. Again, their mortality in the next decade was still higher.

3. Those Congressmen surviving beyond the second decade showed, in turn, a much lower mortality.

4. Those Congressmen whose Public Service was only 6 years or less showed a materially lower mortality than those whose Public Service extended beyond 6 years.

5. The mortality of our modern Congressmen is in general lower than that of their predecessors.

6. This historical improvement in Congressional mortality is greatest for the younger ages—in accordance with the observations on mortality of the general population.

7. For the ages "60 and above" there has been no improvement. This is also in accordance with observations on the mortality of the general population.

Digitized by Google

8. The Congressional mortality during and soon after the Civil War was exceptionally low.

9. This low mortality for the Civil War period was especially marked for the older ages.

10. The last five conclusions (as to historical changes in mortality) are, in general, consonant with the results of the investigations of Mr. E. B. Morris for Yale graduates. (In the case of Yale graduates the mortality at older ages has not only not improved but has grown worse.)

11. Comparing the mortality of Southern, Eastern, and Western Congressmen we find that of the Southerners was the highest and that of the Westerners the lowest, the figures being 100, 93, and 80 per cent. for the South, East, and West respectively.

In the above generalizations no mention has been made of the mortality of Presidents and Vice Presidents. This seems to be high and to show a consistent progressive change for the worse; but the number of lives available is so small, compared with the number of Congressmen considered, as scarcely to justify independent generalizations.

The foregoing generalizations merely give the facts as we find them without reference to their interpretation. The facts may seem to different people capable of different interpretations. My own interpretation is one into which all the facts fit perfectly and which is also in tune with the results of E. B. Morris's investigations for Yale graduates and other studies, such as certain investigations of the United States Census and of the Life Extension Institute. This interpretation is as follows:

(a) Presidents, Vice Presidents, and Congressmen are a select group endowed at the outset with unusual vitality.

(b) The vitality of most of them is severely taxed and materially reduced by the heavy responsibilities and bad personal hygiene incident to a public career.

(c) A minority escape the worst of these conditions and retain the major part of their natural endowment of vitality.

(d) Bad hygiene has been more destructive than the weight of responsibility. This conclusion is suggested by the fact that the Vice Presidents suffer a greater curtailment of life

Digitized by Google

than the Presidents, although Vice Presidents have far less responsibility but spend much more time in Washington social life than the Presidents. It is also suggested by the fact that the Civil War period, with its greater weight of responsibility but more simple living, showed a lower mortality than the periods preceding and following, with less weight of responsibility, but less simple living.

(e) Historically there has been a progressive decrease in those causes of death consisting of *infections* which kill their victims chiefly at the earlier ages, but a progressive change for the worse in those causes of death consisting of the *wear and tear* diseases (or degenerative diseases) which kill their victims chiefly at the later years. These opposite changes in the infectious and degenerative diseases would explain the improvement in mortality at the earlier ages and, on the other hand, the non-improvement (and in the case of the Yale graduates even positive retrogression) in the mortality at the older ages; again they would explain the marked exceptionality of the older ages in the Civil War period; for if it was plain living which explained the low mortality of that period we should expect the most marked influence of this factor to appear, just as it does, at the older ages at which the wear and tear mortality is chiefly registered.

(f) The chief causes of wear and tear on the lives of Congressmen are the same as those applying to modern convivial social life in general; they are the causes emphasized by the Life Extension Institute which, through its Hygiene Reference Board of hygienists and scientists, has described these causes in its recent hand book "How to Live." The chief causes emphasized are indoor and sedentary living, a crouching, cramping posture, late dinners, overeating, especially of nitrogenous foods, such as meat and eggs, overconcentration or lack of bulk in diet, hasty eating, constipation, irregular hours, lack of sleep, worry, and the use of drugs, especially narcotics like alcohol and tobacco. In short, the chief reasons why Congressmen shorten their lives are avoidable, being due, not so much to their work as to the kind of life into which their work leads them.

Digitized by Google

The damage to vitality which we find from public life is not, therefore, put forward as an argument against entering public service; nor will a knowledge of the facts probably tend in that direction, certainly not among those who place public interests above personal interests, as every public man should. But knowledge is power and a knowledge of the facts should enable us to protect our public men or enable them to protect themselves, and to do so in the public interest. This can be accomplished by better general health conditions in the country and its Capital, by better health customs and habits, by better health ideals, and by a more systematic application of the imperfect ideals already existing.

Digitized by Google

THE AGRICULTURAL ELEMENT IN THE POPULATION.*

By Eugene Merritt, A. B., *Scientific and Administrative Assistant, States Relations Service, United States Department of Agriculture.*

It is the function of the agricultural element in the population to produce the raw material to feed and clothe the nation. The Census for the United States does not classify the population according to their source of income, but in many of the European countries such a classification is made. The agricultural element in these countries includes those actively engaged in agriculture, members of their families and others dependent upon them for subsistence, and servants living in the household of their masters. As far as making a study of the agricultural element in the population of the United States is concerned, we have to make it by inference, from a study of those who are gainfully employed in agriculture and of the rural population as defined by the Census. Countries that show not only the number employed in agriculture but the number "dependent" upon agriculture, offer a method of measuring the agricultural element in the population of those countries for which we have only occupation statistics. It appears that the percentage that the male agricultural workers makes of the total male workers is the best index of the relative importance of the agricultural element in the population. This eliminates, or greatly diminishes the influence of the wide variations found in the percentage of the total population that has reached the productive age periods, the differences in the participation of women in the work of the nation and their inclusion in occupation statistics, and also the differences in definitions as to when a person is gainfully employed.

In practically all countries where the number dependent upon agriculture is known, they form a decreasing proportion of the total population. Wherever a comparison of the male

* Paper presented at the annual meeting of the American Statistical Association, Washington, D. C., December 29, 1915.

Digitized by Google

agricultural workers with the total males gainfully employed is available, the agricultural workers form a decreasing proportion of the total. Thus is released to engage in other occupations a corresponding percentage of the total workers. Apparently the principal reasons for this decreasing percentage are that the agricultural element in the population is becoming more efficient and that in the readjustment or changes in the methods of producing and distributing agricultural products, the agricultural people now perform a smaller part of the complete operations than was the case formerly. For example, cheese was manufactured in the home; now it is a factory product. There is a smaller proportion of meat slaughtered and cured on the farm than formerly. Farmers perform a smaller part of the hauling of farm produce to market because the railroads more thoroughly cover the country.

Many persons, in calling attention to the decreasing proportion of the population living in rural districts, feel that this is a national calamity. Indeed if it should happen that an increasing proportion of our people were found on farms it would be a sure sign that our agricultural people were losing their efficiency and should be cause for alarm. If conditions in the United States were similar to those in China there would be between 70 and 75 per cent. of the population engaged in agriculture or dependent on it for their subsistence, whereas in the United States in 1910, only 35 per cent. were so engaged. In other words, the agricultural element in the population of the United States is twice as efficient as the agricultural element in the population of China, to say nothing of the difference in the standards of living of the population of China and that of the United States.

The evidence of the fact that the agricultural element in the population of the United States is becoming more efficient is abundant. The per capita crop production based on total population increased 30 per cent. between 1856 and 1915, while the percentage that the males engaged in agriculture formed of those engaged in all occupations decreased from 50 to 35 per cent. in the last 30 years. In other words, we are producing more crops per capita and use a smaller percentage of our total population for the purpose.

Digitized by Google

As for the meat supply, the per capita production during the last 15 years has remained practically constant or decreased slightly. However, there is a decreased consumption of beef and mutton and an increased consumption of poultry, the consumption of pork remaining practically the same. It is the natural evolution of agricultural nations to consume relatively large quantities of meat in their early development, but as the country becomes more thoroughly settled and the land higher in price, the crops are consumed by the people direct rather than after their transformation into meat. By this means a larger number of persons can be supported on a given acreage.

Another indication of the increased efficiency of our agricultural workers is the fact that the average number of acres of crops cultivated per agricultural worker is increasing and this increase is accompanied by an increased number of horses on farms per agricultural worker, and by an increase in average yields per acre.

If we study that portion of the United States where there has been an actual decrease in rural population, we find that this decrease occurs principally in counties in the North-central, New England, and Middle Atlantic States where there has been a decrease in the number of farms. If we indicate on maps the actual decrease in farms by sizes, the decrease in the number of farms of between 20 and 49 and 50 and 99 acres, will practically explain the decrease in the total number of farms and the consequent decrease in rural population. The farm management studies indicate that on these small-sized farms, man labor, horse labor, and agricultural machinery can not be used efficiently. In other words, economic competition is eliminating the unprofitable sized farms.

It is also interesting to note the type of farm from which the boys and girls are leaving most rapidly. We find that wherever investigations of this nature have been conducted, a smaller percentage of children are found at home on farms between 20 and 100 acres than on those of a larger size.

There is also another change taking place in the Northcentral States that helps to explain the decrease in rural population. The absolute number of foreign born in rural districts of the

Digitized by Google

Northcentral States is decreasing and those of foreign or mixed parentage apparently have reached their maximum numbers. It is a well-known fact that the foreign born have larger families and that the native born have smaller families, and that as the children of the foreign born become more thoroughly Americanized the size of their families decreases. Thus the decrease in rural population in the Northcentral States is due not only to a small number of farm families but also to smaller families.

It is also worthy of notice that the rural population is decreasing in areas where the rural telephone, coöperative agricultural organizations, and coöperative fire insurance companies are to be found in greatest numbers. These are evidences of an efficient and progressive agricultural people.

It is a well-accepted principle of economics that there is a flow of laborers from different classes of occupations and the different markets similar to the flow of money or gold from the different money markets of the world. Anything that tends to upset the relation between the supply and demand of labor in one field of endeavor sets in motion a migration from one part of the country to another or from one country to another. The relationship between the rural and the urban population of the United States is rendered unstable by differences in birth rate, death rate, and by migration.

The following differences in the rural and urban communities influences the number of births. It appears that at all ages a larger percentage of those living in rural districts are married and that the difference is greatest at the youngest ages, that is at the ages of greatest fecundity. The average age at death of females in cities is much lower than that in rural districts, with the result that the urban woman has a shorter married life. That the marriage relationship is more often interrupted by death in urban districts is shown by the greater percentage of widowed persons. When we compare the average number of years married per child born, we find that the average for persons of the same nativity is greater in urban than in rural districts, but that it is lowest for the foreign born and increases as we approach the native born of native parentage. If the population of the United States were made up

Digitized by Google

wholly of persons native born of native parentage, the birth rate in the rural districts would be 40 to 50 per cent. higher than in urban districts, but since the native born of native parentage are in the majority in the rural districts and the foreign born and their children in the urban districts, the relative birth rate appears to be about equal.

There are changes taking place in the relative proportion of the different nativities in the rural districts that have an important bearing upon the future of the rural population. I have already mentioned that in the Northcentral States, the foreign born and their children apparently are either actually decreasing or have reached their maximum, but we find that in the New England and Middle Atlantic States there are less persons native born of native parentage in rural districts today than there were 20 years ago, while those of foreign birth or of foreign and mixed parentage are on the increase. This is evident not only from a study of the relative number of foreign born in the total rural population, but also from a study of the year of immigration, a much larger percentage having come to the New England and Middle Atlantic States in the last ten years than to any other geographic division where the foreign born are found in considerable numbers. It appears that as more intensive agriculture becomes profitable the foreign born begin to take the place of the native born whites. This raises the question as to whether this same evolution will take place throughout the whole United States as our country develops and our entire agricultural system takes on a more intensive nature.

Another factor contributing to the migration from rural to urban districts is the difference in death rates. We find when the death rate is compared age for age, it is higher in the urban than in the rural districts by 10 to 50 per cent. and that the difference is generally greatest at those ages at which the largest number of deaths occur. This has an influence not only on age distribution and migration but upon the length of the working life. For example, the average age at death for all classes of people in 1910, is 35.4 for males in urban districts and 42.4 in rural districts. If we consider that the average age at which our population becomes productive is 15, it means

Digitized by Google

that those dying in the rural districts lived an average productive life of practically 35 per cent. longer than those in urban districts.

The average age for females is 37.4 and 43.3, respectively. Even if the average age at which those in urban and rural districts married were the same, the difference in the age at death would add over five years to the married life of those married. If we make a comparison of those who are native born of native parentage we find that the age at death for urban males is 31.4 and rural males 38.2, and for females 35.2 and 40.7, respectively. This difference in the average age at death does not measure entirely the difference in the average length of the productive life in the two areas, because of the migration from rural districts to cities of persons who are above the average age of the beginning of productivity.

Even if the birth rate for urban and rural districts were the same, the greater death rate in cities would result in a smaller percentage of those born in our cities reaching the productive age period. Even though at the birth there were a proper equilibrium between the potential supply of labor in urban and rural districts, when 15 years of age was reached this equilibrium would be upset, and it would be necessary to transfer some of those living in rural districts to urban districts to bring the supply back to the normal demand.

If we compare the number of persons in our cities between 5 and 14 years of age and 15 and 24 years, we find for native white females of native parentage that there is an excess in the age group 15 to 24 over the number of those 5 to 14. In other words, the migration of young women from rural districts has been sufficiently large not only to overcome the losses by death in urban districts, but to increase from 2 to 20 per cent. the number between 5 and 14 years of age. We find the same condition true for males except that the percentages are not so high and that in three geographic divisions it was not sufficient to overcome the losses by deaths.

This migration from rural districts appears practically to cease at 35 years of age, and the higher death rate in cities accounts for the fact that a larger percentage of those in rural districts survive.

Digitized by Google

Perhaps the most interesting feature in the migration from our rural districts to our cities is that the white women are migrating in greater numbers and that their migration begins at an earlier age. We find that in practically all counties there is an excess of males in the rural districts and when we compare by sex and by age groups the percentage of the total population living in rural districts we find an excess of men at all ages.

However, Negro women form a striking exception to this statement. We find that in the South a majority of the counties have an excess of Negro females in the rural districts, the excess being due principally to an excess of females in the middle age groups. At all other ages the males predominate in the rural districts.

The conditions in the United States are more interesting when compared with the diagrams and statistics relating to conditions in European countries. For example, if we compare the percentage of the total population dependent upon agriculture by ages and sexes, we find that the percentage of females dependent on agriculture is greatest in Austria and Germany up to 60 years of age. Above this year, apparently, the presence of old-age pensions and the classification of people as dependent upon an industry influences the distribution. In the rural districts of England and Wales we find an excess of males up to 15 to 20 years of age, after which there is an excess of females. In the rural districts of Denmark, the excess of females does not begin until 30 years of age is reached.

In European countries where the agriculture is carried on by means of hand labor we find an excess of females in both the agricultural and rural population. The extensive use of agricultural machinery in the United States is probably the cause of the migration of women from rural districts. As long as women are employed in the fields they contribute to the farm income. When an agricultural machine is introduced, it not only takes away the field work of the farm women so that they become of less economic value on the farm, but it also makes them seek remunerative employment elsewhere. For example, if a machine were introduced to pick the cotton

Digitized by Google

of the South, the Negro women would have to seek other fields for employment.

Thus it is evident that the reason for the decreasing percentage of all peoples found in rural districts and the migration of young men and women from our farms, is that as the agricultural element in the population becomes more efficient, a smaller percentage of them is needed on farms and they have to seek employment in the non-agricultural industries.

The higher death rate, age for age, in urban districts depletes the ranks of the workers so that the rural peoples are called upon not only to furnish raw material to feed and clothe the nations, but to fill up the ranks of the city workers and to contribute to the supply of labor demanded by our growing industries.

Digitized by Google

TABLE I.

NUMBER OF PERSONS DEPENDENT UPON AGRICULTURE.

Country.	Year.	Males.		Females.		Both Sexes.	
		Number.	Per Cent. of Total Population.	Number.	Per Cent. of Total Population.	Number.	Per Cent. of Total Population.
Austria.....	1900	6,505,725	50.6	6,941,637	52.2	13,447,362	51.4
	1890	6,333,786	54.2	6,815,414	55.8	13,149,200	55.0
	1880	5,854,945	54.1	6,417,420	56.7	12,272,365	55.4
Hungary.....	1900	6,542,821	68.3	6,518,297	67.4	13,061,118	67.8
	1890	.....	....	.....	....	10,905,444	62.4
Total.....	1900	13,048,546	58.2	13,459,934	58.6	26,508,480	58.4
	1890	.....	....	.....	....	24,054,644	58.2
British India....	1901	.....	....	.....	....	191,691,731	65.1
	1891	.....	....	.....	....	171,735,390	59.8
Bulgaria	1905	1,541,189	74.9	1,548,112	78.2	3,089,301	76.6
	1900	1,394,800	73.0	1,387,469	75.6	2,782,269	74.3
	1893	1,224,534	72.4	1,207,024	74.5	2,431,558	73.4
	1888	1,152,112	71.8	1,163,015	75.1	2,315,127	73.4
Denmark.....	1911	523,046	39.1	480,916	33.9	1,023,962	37.1
	1906	511,814	40.7	479,107	36.0	990,921	38.3
	1901	491,959	41.2	463,803	36.9	955,762	39.0
	1890	445,779	42.1	436,557	39.2	882,336	40.6
	1880	447,360	....	441,571	....	888,931	45.1
	1870	396,621	....	392,114	....	788,735	44.2
	1860	.....	....	.....	....	727,184	45.4
	1845	.....	....	.....	....	614,619	45.5
	1834	.....	....	.....	....	598,558	48.9
Finland.....	1900	775,380	57.8	779,977	56.9	1,555,357	57.3
France.....	1891	8,780,924	46.4	8,654,964	45.1	17,435,888	45.7
	1886	8,824,391	46.7	8,874,011	46.6	17,698,402	46.7
	1881	9,156,873	49.1	9,092,336	48.5	18,249,209	48.8
	1876	9,514,219	51.8	9,454,386	51.0	18,968,605	51.4
	1872	9,231,741	51.3	9,281,584	51.2	18,513,325	51.3
	1866	9,095,515	47.8	9,199,334	48.3	18,294,849	48.1
	1861	.....	....	.....	....	16,030,745	42.9
Germany.....	1907	8,060,282	26.5	9,029,214	28.9	17,089,496	27.7
	1895	8,648,414	34.0	9,420,249	35.7	18,068,663	34.9
	1882	9,047,866	40.8	9,792,952	42.4	18,840,818	41.7
Norway.....	1900	414,479	38.9	440,308	38.1	854,787	38.5
	1891	391,516	41.2	434,735	41.9	826,251	41.5
	1875	436,867	49.8	465,882	50.1	902,749	50.0
Portugal.....	1900	1,646,605	63.5	1,720,594	60.8	3,367,199	62.1
Russia:							
Caucasus....	1897	.....	....	.....	....	7,266,428	78.2
Central Asia..	1897	.....	....	.....	....	6,361,466	82.1
Poland.....	1897	.....	....	.....	....	5,302,850	56.4
Russia Proper.	1897	.....	....	.....	....	69,470,360	74.3
Siberia.....	1897	.....	....	.....	....	4,448,456	77.2
Total.....	1897	.....	....	.....	....	92,849,560	73.9
Servia.....	1900	1,055,300	82.4	1,042,688	86.1	2,097,988	84.2
	1895	975,399	82.2	959,134	85.2	1,934,533	83.7
	1890	900,612	81.1	916,611	87.1	1,817,223	84.1
Sweden.....	1900	1,182,649	47.2	1,161,963	44.2	2,344,612	45.6
	1890	1,204,325	52.0	1,203,461	48.8	2,407,786	50.3
	1880	1,156,678	52.2	1,157,441	49.2	2,314,119	50.7
	1870	.....	....	.....	....	2,286,309	54.8
Switzerland.....	1900	566,682	34.8	501,223	29.7	1,067,905	32.2
	1888	576,019	40.6	530,411	35.4	1,106,430	37.9
	1880	587,895	42.2	550,783	37.9	1,138,678	40.0
	1870	576,594	44.2	555,072	40.7	1,131,666	42.4

Digitized by Google

TABLE II.

NUMBER OF PERSONS ENGAGED IN AGRICULTURE IN VARIOUS COUNTRIES.

Country.	Year.	Males.		Females.		Both Sexes.	
		Number.	Per Cent. of All Occupations.	Number.	Per Cent. of All Occupations.	Number.	Per Cent. of All Occupations.
United States...	1910	10,582,039	35.2	1,806,584	22.4	12,388,623	32.5
	1900	9,272,315	39.0	977,336	18.4	10,249,651	35.3
	1890	7,787,539	41.4	678,824	17.3	8,466,363	37.2
	1880	7,068,058	47.9	594,385	22.5	7,663,043	44.1
Algeria.....	1881	636,078	74.8	91,602	53.7	727,680	71.3
Argentina.....	1895	318,149	28.0	67,174	13.4	385,323	23.6
Australia.....	1901	377,626	29.5	39,029	11.1	416,655	25.6
Austria.....	1900	3,997,660	50.7	4,116,098	70.2	8,113,758	59.1
	1890	4,098,695	55.3	4,295,043	73.9	8,394,638	63.5
	1880	3,427,835	51.7	2,728,830	60.7	6,156,665	55.4
	1870	.....	....	.....	....	7,506,395	62.6
Hungary.....	1900	4,187,590	68.6	1,819,707	70.4	6,007,297	69.1
	1890	3,913,001	71.8	1,469,675	67.1	5,382,676	70.5
Total.....	1900	8,185,250	58.5	5,935,805	70.3	14,121,055	63.0
	1890	8,011,696	62.3	5,765,618	72.0	13,777,314	66.1
Belgium.....	1900	533,665	23.6	163,707	17.6	697,372	21.9
	1890	489,545	24.6	159,707	19.1	649,252	23.0
	1880	542,006	30.8	265,458	28.8	807,464	30.1
	1856	709,214	48.9	352,901	42.7	1,062,115	46.7
	1846	550,567	42.4	356,008	47.0	906,575	44.0
Bolivia.....	1900	.....	....	.....	....	564,009	43.5
British India....	1901	63,026,365	67.3	27,867,210	66.5	90,893,575	67.1
	1881	52,029,098	64.1	19,169,870	50.7	71,198,968	59.9
British North Borneo...	1901	.....	....	.....	....	32,892	64.2
Bulgaria	1905	895,206	73.3	837,406	94.9	1,732,612	82.4
	1900	798,928	71.5	730,870	93.1	1,529,798	80.4
	1893	652,723	71.4	657,324	73.9	1,310,047	72.6
	1888	621,196	72.7	677,445	82.2	1,298,641	77.4
	1881	415,316	70.8	349,552	90.7	764,868	78.7
Canada.....	1901	707,997	45.4	8,940	3.7	716,937	39.9
Ceylon.....	1901	745,074	65.0	318,551	65.4	1,063,625	65.1
Chile.....	1907	448,546	50.3	21,877	6.2	470,423	37.7
Cuba.....	1907	364,821	52.2	3,110	4.2	367,921	47.6
	1899	.....	....	.....	....	299,197	48.1
Cyprus.....	1901	33,611	62.8	2,757	20.8	36,368	54.5
Denmark.....	1911	386,016	45.7	110,169	28.5	496,185	40.3
	1901	366,860	43.4	149,760	41.5	516,620	46.2
Egypt.....	1907	2,258,005	67.2	57,144	33.3	2,315,149	65.6
	1897	2,049,643	64.5	.....	....	2,049,643	63.1
Federated Malay States....	1901	115,027	28.2	52,324	82.7	167,351	35.5
Finland.....	1900	321,538	51.4	102,008	39.6	423,546	48.0
Formosa.....	1905	763,456	70.6	263,664	82.4	1,027,120	73.3
France.....	1905	5,452,392	41.9	3,324,661	43.2	8,777,053	42.4
	1901	5,454,940	42.3	2,654,316	39.0	8,109,256	41.1
	1896	5,611,628	44.7	2,747,709	42.9	8,359,337	44.1
	1891	4,694,714	42.4	1,840,885	35.9	6,535,599	40.3
	1886	4,777,729	43.2	2,138,236	37.8	6,915,965	41.4
	1881	4,603,412	41.1	1,650,411	30.1	6,253,823	37.5
	1876	5,088,337	48.1	1,548,113	30.3	6,636,450	42.3
	1872	4,579,838	....	1,298,085	....	5,877,923	40.2
	1866	4,629,752	44.1	1,568,234	33.8	6,197,986	40.9
Germany.....	1907	5,146,723	27.7	4,585,749	48.3	9,732,472	34.6
	1895	5,410,205	34.8	2,745,840	41.7	8,156,045	36.9
	1882	5,588,892	41.7	2,531,626	45.7	8,120,518	42.8
Greece.....	1907	321,120	47.3	6,972	12.2	328,092	44.6
Grenada	1901	8,816	57.1	7,722	49.7	16,538	53.4
Italy.....	1901	6,370,277	57.9	3,196,063	60.5	9,566,340	58.8
	1882	5,396,901	53.0	3,094,272	47.2	8,491,173	50.7
	1872	5,537,427	57.5	3,027,466	53.0	8,564,893	55.9
	1862	4,869,421	58.1	2,839,210	51.2	7,708,631	55.4
Jamaica.....	1911	.....	....	.....	....	271,493	66.1
	1891	.....	....	.....	....	271,296	72.6
	1881	.....	....	.....	....	208,587	71.7
Malta and Gozo.	1901	10,235	12.3	3,613	15.8	13,848	13.9
Mauritius.....	1901	72,493	57.1	5,989	38.0	78,482	55.0

Digitized by Google

TABLE II—*Concluded.*

Country.	Year.	Males.		Females.		Both Sexes.	
		Number.	Per Cent. of all Occupations.	Number.	Per Cent. of All Occupations.	Number.	Per Cent. of all Occupations.
Netherlands	1899	490,694	32.9	79,584	18.4	570,278	29.6
	1889	451,756	34.8	72,868	20.6	524,624	31.7
	1859	217,935	19.0	35,698	7.3	253,633	15.5
	1849	229,422	21.7	95,349	18.5	324,771	20.6
New Zealand ...	1911	103,644	28.5	7,472	8.3	111,116	24.5
	1906	93,504	28.9	3,453	4.6	96,957	24.3
	1901	85,323	31.1	3,899	5.9	89,222	26.2
	1896	80,194	33.4	3,106	5.9	83,300	28.4
	1891	65,950	31.8	2,657	5.9	68,607	27.1
	1886	64,259	34.1	919	2.8	65,178	29.5
	1881	53,769	33.0	678	2.7	54,447	29.0
Norway.....	1910	.....	....	.....	....	307,528	33.4
	1900	237,705	39.6	71,311	25.3	309,016	35.0
	1891	226,688	42.7	84,890	34.7	311,578	40.2
	1875	170,995	37.9	15,142	7.9	186,137	28.9
Philippine Is....	1903	1,163,777	57.8	90,286	8.8	1,254,063	41.3
Porto Rico	1899	196,893	73.3	1,868	3.9	198,761	62.8
Portugal.....	1900	1,127,268	65.3	380,293	52.0	1,507,561	61.4
Russia:							
Russia Proper.	1897	11,464,216	60.4	1,671,399	38.0	13,135,615	56.2
Poland.....	1897	1,074,914	48.0	218,706	37.0	1,293,620	45.7
Caucasia.....	1897	1,269,375	65.2	84,059	40.8	1,353,434	62.9
Central Asia .	1897	1,251,416	73.7	32,329	20.8	1,283,745	69.3
Siberia.....	1897	841,549	63.5	72,808	38.4	914,357	60.4
Total.....	1897	15,901,470	60.7	2,079,301	37.5	17,980,771	56.7
St. Lucia.....	1901	.....	....	.....	....	15,796	54.1
Servia.....	1900	311,700	65.5	13,524	50.5	325,224	64.7
	1895	.....	....	.....	....	299,060	67.9
	1890	.....	....	.....	....	288,539	70.6
Sierra Leone....	1901	8,705	28.7	4,544	21.7	13,249	25.9
Spain.....	1900	3,741,730	58.1	775,270	51.8	4,517,000	56.9
	1887	4,033,391	72.0	821,351	57.6	4,854,742	69.1
Sweden.....	1900	761,016	52.4	333,264	53.8	1,094,280	52.8
	1890	759,925	58.4	330,266	58.6	1,090,191	58.5
	1880	728,944	59.1	329,054	60.4	1,057,998	59.5
Switzerland.....	1900	392,971	37.1	80,326	16.1	473,297	30.4
	1888	388,467	44.6	92,566	21.3	481,033	36.8
	1880	398,873	44.5	147,589	29.1	546,462	38.9
	1870	388,297	47.5	159,944	34.3	548,241	42.7
Trinidad and Tabago ..	1901	51,744	54.7	25,765	39.3	77,509	48.4
Union of South Africa:							
Cape of Good Hope	1904	524,228	64.9	485,246	81.7	1,009,474	72.0
	1891	383,612	75.0	288,846	79.2	672,458	76.8
Natal.....	1904	33,716	36.9	6,416	51.2	40,132	38.6
Orange River Colony...	1904	91,986	60.2	61,028	50.8	153,014	56.0
Transvaal....	1904	213,293	44.5	294,367	80.3	507,660	60.0
Total.....	1904	863,223	56.3	847,057	77.5	1,710,280	65.1
England and Wales....	1901	1,158,940	11.4	38,982	0.9	1,197,922	8.4
	1891	1,234,101	14.0	51,045	1.3	1,285,146	10.1
	1881	1,288,328	16.6	64,216	1.9	1,352,544	12.1
	1871	1,391,630	21.2	186,101	5.7	1,577,731	16.0
Scotland.....	1901	176,816	12.7	28,191	4.8	205,007	10.3
	1891	183,902	15.3	29,158	5.4	213,060	12.2
	1881	188,533	17.3	51,647	10.7	240,170	15.2
	1871	187,947	19.2	50,464	10.3	238,411	16.2
Ireland.....	1901	774,056	55.2	85,469	15.6	859,525	44.1
	1891	831,753	57.2	90,967	15.4	922,720	45.1
	1881	885,518	58.1	95,838	15.1	981,356	45.4
	1871	891,890	53.6	170,118	19.3	1,062,008	41.7
	1861	904,748	49.5	84,181	10.0	988,929	37.0
United Kingdom	1901	2,109,812	16.3	152,642	2.9	2,262,454	12.4
	1891	2,249,756	19.6	171,170	3.4	2,420,926	14.6
	1881	2,362,369	22.8	211,701	4.7	2,574,070	17.3
	1871	2,471,467	26.8	406,683	8.8	2,878,150	20.8

Digitized by Google

TABLE III.

PERCENTAGE OF TOTAL RURAL AND URBAN POPULATION OF THE UNITED STATES, CLASSIFIED BY NATIVITY.

Nativity and Geographic Division.	Rural.			Urban.		
	1910.	1900.	1890.	1910.	1900.	1890.
Native Parentage.						
New England	69.8	73.0	77.6	33.9	37.8	43.6
Middle Atlantic	67.0	70.6	71.3	34.4	35.8	35.8
East North Central	66.4	64.9	64.3	41.7	38.8	36.8
West North Central	58.5	57.5	59.3	51.2	47.8	45.6
South Atlantic	62.2	60.8	59.9	54.2	50.0	45.9
East South Central	67.2	65.3	65.0	54.4	47.6	44.1
West South Central	67.7	64.4	61.4	58.4	47.4	41.7
Mountain	57.8	52.9	50.4	51.9	47.2	47.6
Pacific	54.8	54.8	54.6	46.9	40.6	36.8
Total, United States	64.1	63.2	63.3	41.9	40.2	39.7
Foreign or Mixed Parentage.						
New England	17.0	14.9	11.5	34.2	31.6	26.6
Middle Atlantic	17.6	16.6	15.5	33.6	34.8	33.3
East North Central	22.4	23.0	21.3	33.0	35.8	33.2
West North Central	27.4	26.7	22.6	28.2	30.4	27.5
South Atlantic	1.4	1.4	1.4	10.1	12.2	13.0
East South Central	1.0	1.2	1.4	9.5	13.2	14.8
West South Central	5.4	5.5	4.3	12.1	16.9	18.9
Mountain	21.2	24.1	21.2	27.3	30.2	25.0
Pacific	22.4	23.4	19.6	27.2	31.4	28.2
Total, United States	13.3	13.6	12.4	29.0	30.9	28.7
Foreign Born.						
New England	12.6	11.5	10.4	30.7	29.3	28.6
Middle Atlantic	13.9	11.2	11.7	29.5	26.8	28.7
East North Central	10.2	11.1	13.2	22.8	22.9	27.5
West North Central	12.6	13.9	15.8	16.3	17.0	21.8
South Atlantic	1.1	0.8	0.8	6.2	6.5	8.3
East South Central	0.4	0.5	0.7	3.7	5.0	7.6
West South Central	3.1	3.1	3.2	7.0	8.8	12.3
Mountain	15.7	15.9	18.6	18.3	19.9	24.0
Pacific	18.4	16.5	18.7	22.2	23.1	28.0
Total, United States	7.5	7.5	8.7	22.6	22.2	24.8
Colored.						
New England	0.6	0.6	0.5	1.2	1.3	1.2
Middle Atlantic	1.5	1.6	1.5	2.5	2.6	2.2
East North Central	1.0	1.0	1.2	2.5	2.5	2.5
West North Central	1.5	1.9	2.3	4.3	4.8	5.1
South Atlantic	35.3	37.0	37.9	29.5	31.3	32.8
East South Central	31.4	33.0	32.9	32.4	34.2	33.5
West South Central	23.8	27.0	31.1	22.5	26.9	27.1
Mountain	5.3	7.1	9.8	2.5	2.7	3.4
Pacific	4.4	5.3	7.1	3.7	4.9	7.0
Total, United States	15.1	15.7	15.6	6.5	6.7	6.8

Digitized by Google

TABLE IV.

CHANGES IN THE RURAL AND URBAN POPULATION IN 1910 AND 1900 COMPARED WITH 1890.

(1890=100 per cent.)

Census Year and Grand Divisioh.	Native Parentage.		Foreign or Mixed Parentage.		Foreign Born.		Colored.		Total.	
	Rural.	Urban.	Rural.	Urban.	Rural.	Urban.	Rural.	Urban.	Rural.	Urban.
1910.										
New England	86.7	119.0	142.4	196.7	116.1	164.6	118.4	156.3	96.3	153.2
Middle Atlantic	97.8	179.9	118.2	188.6	123.7	192.0	109.0	220.4	104.1	187.1
East North Central	106.4	213.8	108.1	187.7	79.4	156.4	88.7	186.9	103.0	188.7
West North Central	115.5	188.5	142.3	171.6	94.0	125.4	76.6	143.0	117.2	167.8
South Atlantic	132.6	211.2	128.8	139.7	166.9	134.0	119.0	160.6	127.7	178.9
East South Central	126.0	238.0	85.0	124.4	74.2	92.7	116.0	185.8	121.8	192.6
West South Central	197.3	382.8	211.3	174.1	164.2	155.4	129.5	228.0	169.6	273.4
Mountain	225.4	290.8	197.0	290.9	165.2	202.7	106.1	195.0	196.4	266.4
Pacific	167.1	378.9	190.5	286.4	164.0	235.4	103.5	157.4	166.6	297.1
Total, United States	124.3	197.8	131.3	189.5	106.4	171.0	118.2	180.5	122.7	187.6
1900.										
New England	92.7	109.0	127.5	148.8	108.6	128.4	120.8	138.8	98.5	125.5
Middle Atlantic	99.0	137.7	107.4	143.6	96.1	127.9	108.1	163.1	100.1	137.4
East North Central	105.5	149.1	112.8	152.7	87.7	117.9	93.5	145.4	104.6	141.6
West North Central	108.2	133.8	132.4	141.0	98.7	99.5	91.1	120.0	111.7	127.6
South Atlantic	116.8	140.8	120.9	120.9	106.1	102.1	112.6	123.1	115.2	129.2
East South Central	114.8	149.5	105.2	123.5	83.0	91.8	114.4	141.0	114.3	138.4
West South Central	142.8	167.9	171.7	131.4	132.3	105.8	118.0	146.8	136.0	147.7
Mountain	138.7	151.1	150.2	183.5	113.1	126.2	94.9	121.7	132.0	152.2
Pacific	119.7	154.6	142.6	155.6	105.2	115.3	87.3	98.7	119.1	140.0
Total, United States	112.2	137.2	122.9	146.0	97.3	121.1	112.8	134.4	112.4	135.5

Digitized by Google

TABLE V.

AVERAGE AGE AT DEATH BY COLOR AND NATIVITY.

Color and General Nativity.	1900.				1910.			
	Cities in Registration States.(a)		Rural Part of Registration States.(a)		Cities in Registration States.(a)		Rural Part of Registration States.(a)	
	All Ages.	15 Years and Over.	All Ages.	15 Years and Over.	All Ages.	15 Years and Over.	All Ages.	15 Years and Over.
Aggregate	32.4	51.5	44.7	59.5	36.3	53.0	42.8	59.1
Males	31.0	50.5	44.3	59.8	35.4	51.8	42.4	58.6
Females	33.9	52.6	45.1	59.1	37.4	54.4	43.3	59.7
White	32.6	51.8	44.9	59.6	36.6	53.4	43.1	59.4
Males	31.1	50.7	44.5	59.9	35.6	52.1	42.6	58.8
Females	34.2	52.9	45.3	59.2	37.8	54.9	43.7	60.0
Native	22.7	48.8	41.1	58.9	27.9	50.7	38.5	58.5
Males	21.5	47.6	40.4	59.4	26.9	49.7	37.8	58.5
Females	24.1	50.0	41.9	58.5	29.0	51.8	39.3	58.6
Both parents native	31.2	55.8	44.6	60.6	33.2	54.1	39.4	59.2
Males	29.1	54.6	44.2	61.1	31.4	52.9	38.2	59.3
Females	33.3	57.0	45.1	60.0	35.2	55.3	40.7	59.1
One or both parents foreign	13.0	36.9	20.1	43.0	19.5	43.3	23.9	48.2
Males	12.7	36.6	19.5	43.3	19.3	42.8	23 8	48.3
Females	13.4	37.2	20.8	42.6	19.7	43.9	24.1	48.1
Foreign	53.4	54.7	60.2	61.4	55.7	56.8	61.0	62.0
Males	52.6	53.9	60.5	61.6	54.2	55.2	59.3	60.2
Females	54.3	55.6	59.8	61.1	57.5	58.8	63.4	64.6
Colored	26.6	45.1	31.4	51.0	31.1	45.1	32.4	49.2
Males	26.3	44.9	32.3	52.9	31.2	44.8	33.6	49.7
Females	26.9	45.3	30.6	49.1	31.0	45.5	31.0	48.6

(a) Registration states in 1900 include Connecticut, District of Columbia, Indiana, Maine, Massachusetts, Michigan, New Hampshire, New Jersey, New York, Rhode Island and Vermont and in 1910 also California, Colorado, Maryland, Minnesota, Montana, North Carolina, Ohio, Pennsylvania, Utah, Washington, and Wisconsin.

Digitized by Google

TABLE VI.

PERCENTAGE OF TOTAL NATIVE WHITE POPULATION OF SAME AGE AND SEX LIVING IN RURAL DISTRICTS IN 1910.

Age Group.	New England.		Middle Atlantic.		East North Central.		West North Central.		South Atlantic.	
	Male.	Female.	Male.	Female.	Male.	Female.	Male.	Female.	Male.	Female.
Under 5 years...	16.4	16.2	30.3	30.3	51.2	50.9	74.0	73.8	79.4	79.1
5–9.....	17.8	17.6	32.7	32.2	54.2	53.6	74.8	74.4	79.5	79.1
10–14.....	18.6	17.7	33.0	32.1	54.7	53 2	74.3	72.9	79.1	77.9
15–19.....	18.6	17.1	32.7	30.1	52.4	49.1	71.4	67.6	76.3	74.3
20–24.....	18.2	16.5	30.8	28.3	48.0	44.8	66.0	62.4	71.8	70.2
25–34.....	19.7	18.0	31.3	29.9	47.4	45.8	63.7	62.1	69.7	68.8
35–44.....	22.1	20.1	34.6	32.7	51.3	48.9	64.7	63.1	70.2	68.4
45–64.....	28.2	25.2	43.0	39.7	58.2	54.2	67.3	63.8	74.5	70.8
65 and over.....	38.1	32.7	57.5	51.0	68.4	62.0	71.8	67.1	78.9	74.4
Total.....	21.2	19.7	34.5	32.8	52.8	50.4	69.4	67.4	75.2	73.6

Age Group.	East South Central.		West South Central.		Mountain.		Pacific.		United States.	
	Male.	Female.	Male.	Female.	Male.	Female.	Male.	Female.	Male.	Female.
Under 5 years...	86.9	86.7	83.6	83.4	69.3	69.3	49.9	49.8	58.0	57.6
5–9.....	86.7	86.2	83.1	82.8	68.9	68.4	51.8	51.0	59.8	59.2
10–14.....	85.9	84.8	82.5	81.4	67.6	65.9	51.0	49.3	59.4	57.9
15–19.....	83.8	82.1	81.2	78.7	66.1	61.4	47.4	42.9	57.5	54.4
20–24.....	79.9	78.8	76 0	74.2	64.9	58.2	42.6	37.0	54.0	50.8
25–34.....	77.8	77.2	78 4	72.7	62.2	59.5	39.7	36.5	52.5	50.5
35–44.....	77.9	76.9	73.3	72.5	62.6	57.0	42.0	37.6	53.7	50.8
45–64.....	81.7	78 9	77 4	74.0	65.4	57.5	47.3	39.6	59.2	54.5
65 and over.....	86.8	82.9	80.0	75.6	67.5	56.1	54.3	41.1	66.4	59.8
Total.....	82.8	81.6	78.9	77.8	65.6	62.0	45.9	42.2	57.0	54.6

Digitized by Google

TABLE VII.

PERCENTAGE OF MALES IN THE TOTAL RURAL AND TOTAL URBAN NATIVE WHITE POPULATION OF THE SAME AGE IN 1910.

Age Group.	New England.		Middle Atlantic.		East North Central.		West North Central.		South Atlantic.	
	Rural.	Urban.	Rural.	Urban.	Rural.	Urban.	Rural.	Urban.	Rural.	Urban.
Under 5 years...	50.6	50.4	50.6	50.5	50.9	50.5	50.7	50.5	51.1	50.6
5–9.....	50.6	50.3	50.6	50.2	50.8	50.2	50.7	50.2	50.9	50.3
10–14.....	51.5	49.9	51.1	49.8	51.1	49.7	51.1	49.3	51.4	49.5
15–19.....	52.1	49.4	51.5	48.5	51.6	48.2	51.6	47.1	50.6	47.9
20–24.....	51.0	47.9	50.4	47.4	50.9	47.6	51.8	47.9	49.6	47.8
25–34.....	50.5	47.7	49.9	48.3	50.5	48.9	51.9	50.2	50.0	48.8
35–44.....	51.4	48.4	50.8	48.7	51.7	49.4	53.3	50.6	51.6	49.4
45–64.....	52.2	48.3	51.4	48.0	53.2	49.1	55.2	51.4	53.1	48.4
65 years and over	49.3	43.4	49.5	43.0	52.7	45.6	54.7	49.1	49.2	42.9
Total.....	51.1	48.8	50.7	48.8	51.5	49.2	52.1	49.8	51.0	48.9

Age Group.	East South Central.		West South Central.		Mountain.		Pacific.		United States.	
	Rural.	Urban.	Rural.	Urban.	Rural.	Urban.	Rural.	Urban.	Rural.	Urban.
Under 5 years...	51.1	50.7	51.0	50.6	50.6	50.6	51.0	50.8	50.9	50.5
5–9.....	50.9	50.0	50.9	50.3	50.8	50.2	51.0	50.3	50.8	50.2
10–14.....	51.4	49.3	51.3	49.5	51.5	49.6	51.4	49.7	51.2	49.7
15–19.....	50.7	47.7	51.1	47.2	52.6	47.5	53.0	48.5	51.4	48.3
20–24.....	49.2	47.5	50.8	48.4	56.7	49.6	56.3	50.5	51.0	47.9
25–34.....	50.0	49.1	52.0	51.2	57.7	52.9	56.5	53.2	51.3	49.3
35–44.....	51.4	50.0	54.1	53.1	59.3	53.6	57.7	53.2	52.6	49.8
45–64.....	53.8	49.4	57.6	53.0	63.4	55.4	60.5	52.8	54.3	49.5
65 years and over	51.4	43.8	54.8	48.0	63.5	51.7	62.9	49.9	52.2	45.1
Total.....	51.1	49.1	52.2	50.6	55.4	51.6	55.3	51.6	51.7	49.3

Digitized by Google

THE MOTHER AND INFANT MORTALITY.

By Henry H. Hibbs, Jr., *Sometime Fellow in Research, Boston School for Social Workers.*

Even the most hasty examination of almost any report on vital statistics will show how widely rates of infant mortality vary in different countries, states, cities, and other communities. This is a phenomenon of great significance to students of infant mortality because it brings us face to face with the question of causes. Why is the infant mortality rate lower in one country than in another, in one city of the same country than in another, in one ward or locality of the same city than in another? Thus, the wide variation in infant mortality in different areas and localities immediately challenges investigation into the relation of housing and living conditions, of sanitation and congestion, of rates of wages, and of social conditions in general, to the proportion of the infants born to those who die in such areas. In the same way any similar variation found to exist, when the infant death rates of a large number of families are compared, will suggest an inquiry into the relationship existing between the proportion of infant deaths in families and the age of the mother, the size of the family, the character and intelligence of the parents, and domestic conditions in general. It is with one phase of this aspect of the problem that this article will deal—the relation of the mother, her age, her character, her intelligence, and her knowledge of infant hygiene, to infant mortality and to her baby's health.

It is obvious that this variation in the proportion of infant deaths in families cannot be shown by the use of the usual infant mortality rate based on the proportion of deaths to 1,000 births because the number of both births and deaths in any family is too small for such a comparison. The only feasible method is the very indirect one of comparing the number of infant deaths that have previously occurred in families with the infant mortality rate for such families in any given year. If the families in which a large proportion of infant deaths have previously occurred contribute a larger number of deaths to the total infant mortality of the year in

Digitized by Google

question than those families in which no previous infant deaths have occurred, it will follow that in the former group of families infants tend to die in larger proportions than in the latter; or, in other words, that the proportion of infant deaths to births is year after year larger in some families than in others. This was the method followed in an investigation of infant mortality in Wards 6, 8, 13, and 17 of Boston by the Research Department of the Boston School for Social Workers.* The results of this phase of the inquiry are set forth in the following table where the mortality rate for infants born in 1910 and classified according to the number of infant deaths that occurred among the children previously born to their mothers is given—the number of previous children born into the family being taken into account by an additional classification of the mothers according to the number of their previous pregnancies:

TABLE SHOWING (A) NUMBER OF BIRTHS AND (B) DEATHS PER 1,000 BIRTHS IN 1910 IN FAMILIES VISITED IN BOSTON CLASSIFIED ACCORDING TO THE NUMBER OF DEATHS DURING INFANCY WHICH OCCURRED AMONG CHILDREN BORN IN PREVIOUS YEARS AND THE NUMBER OF THE PREVIOUS PREGNANCIES OF THE MOTHERS. (a)

Families Classified According to the Number of Deaths During Infancy Which Occurred Among Children Born Previous to 1910.	Families Classified According to the Number of the Mother's Pregnancies Occuring Previous to 1910,			
	Total.	1, 2, or 3	4, 5, or 6	Over 6
(A) NUMBER OF BIRTHS IN 1910 IN:				
All families (b)	1671	961	487	223
Families with no previous infant deaths	1192	807	297	88
Families with one or more previous infant deaths	479	154	106	135
Families with—				
One previous infant death	331	134	118	79
Two previous infant deaths	115	19	62	34
Three or more previous infant deaths	33	1	10	22
(B) DEATHS PER 1,000 BIRTHS IN 1910 IN:				
All families	134	126	126	146
Families with no previous infant deaths	117	114	125	125
Families with one or more previous infant deaths	173	188	179	148
Families with—				
One previous infant death	154	194	153	89
Two previous infant deaths	191	158	210	177
Three or more previous infant deaths	303	...	300	318

(a) This table is intended to show the relationship between the infant death rate in 1910 in the families visited and the infant death rate during previous years in the same families. Since it is not possible to compile a true infant mortality rate based on the ratio of deaths to 1,000 births for so small a unit as the family this double classification, according to the number of previous infant deaths and the number of the mother's previous pregnancies—this last being a rough approximation of the number of births—has been substituted.

(b) Information was not obtained in 21 instances. The 371 infants who were born during their mother's first pregnancy are omitted from this table.

* The writer is indebted to Dr. J. R. Brackett, director of the School for Social Workers, for the use of these figures which have not been previously published. The investigation was made in the academic years 1910–11 and 1911–12. During the second of these years the field work was carried on under the direction of the writer supervised by the director of the Research Department, Dr. T. W. Glocker.

Digitized by Google

An examination of this table shows that the rate of mortality for infants born to mothers, none of whose previous children have died during infancy, is strikingly less than the rate for those whose mothers have already had one or more infant deaths. Moreover, the rate rises as the number of previous infant deaths in the family increases and varies in the same manner when the infants are also classified according to the number of their mothers previous pregnancies.* The conclusion that infant mortality rates for families vary as widely as those for states and cities appears, therefore, to be justified. The table also shows that the rate of infant mortality in any community is determined, not by the relatively uniform occurrence of deaths in all families, but by their relatively frequent occurrence in certain especially unfortunate families, and that the high mortality rate for the period of infancy is to be laid at the door of these relatively few families where the proportion of infant deaths to children born is especially large.

This, however, is merely another way of saying that the proportion of infant deaths to births in families varies in the same way that it does in other divisions of the population. Having shown this—and it would probably have been admitted *a priori* by many— it will be necessary to study at length the nature of some of the domestic and social conditions which cause the regular occurrence of a high ratio of infant deaths to births in certain families, while others escape with a smaller proportion of deaths or none at all; and to ascertain, if possible, why in any district parents may be found who have successfully reared every member of a large family, while side by side with them are other families in which numerous infant deaths have occurred. Among these "domestic factors" of infant mortality one of the most important is the mother, her age, her health, her character and intelligence, and her devotion to the child.

The Influence of the Age of the Mother at the Birth of the Child. The relation of the age of the mother at the birth of her child to the rate of infant mortality has been frequently commented upon, though not always made the subject of careful study. It is often asserted on the one hand that young mothers, because their strength has not been depleted

* The one exception in the case of the mothers with one, two, or three previous pregnancies and two previous infant deaths is probably due to chance as there were only nineteen births and three deaths included in the group.

Digitized by Google

by previous childbearing or by the cares of a family, tend, other things being equal, to give birth to healthy children and that the rate of mortality among their infants is, therefore, very low. On the other hand, it is probably as often claimed that young mothers are too immature to give birth to healthy and well formed children and, in addition, often incapable through the lack of previous experience of giving them the needed care after birth, as a consequence of which the mortality among their infants is relatively high. The following table summarizes the data bearing on this aspect of the subject which have been collected in five recent investigations:

MORTALITY RATE PER 1,000 BIRTHS AMONG INFANTS INCLUDED IN FIVE INVESTIGATIONS, CLASSIFIED ACCORDING TO THE AGE OF THEIR MOTHERS.

Age of the Mother.	Number of Births.	Infant Mortality Rate.
Boston (Mass.) Investigation. (a)		
All ages	2,025	125
Under 21 years	145	90
21 to 25 years	559	109
26 to 30 years	573	131
31 to 35 years	440	132
36 to 40 years	241	149
Over 40 years	67	164
Fall River (Mass.) Investigation. (b)		
All ages	746	202
Under 20 years	29	103
20 to 29 years	386	189
30 to 39 years	257	206
40 and over	36	222
Unknown	38	—
Johnstown (Pa.) Investigation. (c)		
All ages	1,463	134
Under 20 years	95	137
20 to 24 years	454	121
25 to 29 years	391	143
30 to 39 years	449	136
40 years and over	74	149
Ewart's Investigation (England). (d)		
Under 19 years	152	171
20 to 24 years	536	132
25 to 29 years	396	166
30 to 34 years	316	170
35 to 39 years	150	220
Over 40 years	36	330
Birmingham (Eng.) Investigation. (e)		
All ages	3,773	176
Under 25 years	936	207
25 to 35 years	1,982	167
35 years and over	855	163

(a) No information was obtained in 38 instances. For source of data see first note.

(b) Dublin: Infant Mortality in Fall River, Mass.—A Survey of the Mortality among 833 Infants Born in June, July, and August, 1913. *Quarterly Publications* of the American Statistical Association, New series, No. 110, June, 1915, p. 515.

(c) U. S. Children's Bureau: Infant Mortality: Johnstown, Pa., Washington, 1915, p. 35.

(d) R. J. Ewart, M. D.: The Aristocracy of Infancy and the Conditions of Its Birth. The *Eugenics Review*, Vol. III, p. 166. The writer does not state where his data were collected but from a later article in the same journal it is to be inferred that they were collected in Middlesborough, England.

(e) Health Department of the City of Birmingham, England: Report on Industrial Employment of Married Women and Infant Mortality, 1911 (p. 11) and 1912 (p. 11).

Digitized by Google

Interpretation of the data presented in this table is somewhat difficult because of the complex factors involved and because the results of the five investigations differ considerably. In the first two, Boston and Fall River, the rate of mortality was lowest for the infants born to the very young mothers under 20 and 21 years of age. In both also the rate increased regularly with the age of the mother after the twentieth year, the highest rate occurring among the infants whose mothers were over 40. In two others, Johnstown and Ewart's, the rate was higher for the children of the very young mothers under 19 or 20 than for those whose mothers were between the ages of 20–30 and 30–35 years. These two investigations agree with the first two, however, in showing that the rate is highest of all among the infants whose mothers were 40 years of age or over. Finally, when the figures for Birmingham are examined, exactly the opposite tendencies are revealed, the rate for the older mothers being lower than that for the younger.

There seems to be a general agreement in the results of the first four of these investigations in showing that the chances of survival are greater for infants born to mothers between the ages of 20–25, and 35–40 than for those born to mothers over 40. The results for Birmingham do not share in this agreement but in interpreting these figures the fact must be considered that no separate data were given for the very young mothers under 20 or the very old mothers over 40 years of age. The results of the five investigations are not in sufficient accord to justify any conclusion as to the rate of mortality among the infants of very young mothers. This question must, therefore, be left in doubt—at least as far as the available statistics are concerned.

On the other hand, in drawing conclusions from these figures the mere fact that the results of the five investigations differ is in itself significant. It indicates, in the writer's opinion, that the age of the mother is not a factor in infant mortality of primary importance. The figures do not show that mothers can obtain the knowledge necessary for bearing and rearing babies only from experience acquired with age. In fact, except for the very young mothers, these figures seem to show that, as far as infants are concerned, one may fairly

Digitized by Google

expect as high a degree of efficiency in motherhood from the young and inexperienced mothers as from the older—a conclusion which from the standpoint of prevention it is not necessary to point out.

The Influence of the Character and Intelligence of the Mother. But if the age of the mother is not a factor in infant mortality of primary importance, her health, her character, and her intelligence certainly are. It is difficult for the careful student of this problem to avoid the conclusion that the real underlying factor of infant mortality and the chief consideration in the health and welfare of babies is the strength, character, health, and intelligence of the mother. For instance, it is the mother's health and strength that determines whether it is physically possible for the baby to start life with the tremendous advantage that comes from breast feeding and it is her character and intelligence that largely determine how the child shall be fed when either method is possible Again, in the recent campaign that has arisen for the control of the influence of prenatal conditions on infant mortality it is evident that the mother is the all-important consideration. Certainly, while the child is developing in the mother's womb her health and strength, the care she takes of herself, and her general standards of hygiene are among the most important considerations in determining whether or not the baby after its birth will get a fair start in life or die in a few weeks or months from prenatal causes.*

But even after the physical connection between mother and child is severed at birth the infant is little less dependent upon its mother, for she determines to a very large extent the character of everything with which it comes into contant—food, sunlight, cleanliness, and even the kind of air it breathes. In both the Boston and Johnstown investigations a close relationship between bad housing conditions and infant mortality was found to exist, but it should not be forgotten in interpreting these figures that the closeness of this relationship is also to a large extent determined by the mother.† From even the briefest investigation it will be evident that even in the

* See the writer's article on "The Influence of Prenatal Conditions on Infant Mortality," Proceedings of the Southern Sociological Congress, 1915.

† See the writer's article on "Infant Mortality and Urban Housing, and Living Conditions," *Journal of Sociologic Medicine* (American Academy of Medicine), October, 1915.

Digitized by Google

poorest districts, where the housing and living conditions are exceedingly bad, many instances in which parents have reared large families without the occurrence of a single death are to be found, while side by side with these in the same neighborhood or even in the same house are to be found other families in which numerous deaths of infants have occurred. Such instances as these, where other differences between the families concerned appear to be so small, seem logically to be the result not of poverty, or the cleanly or sanitary condition of the home, or even of the physical strength of the parents, but of the amount of intelligent attention and care which the mother bestows upon her baby. In such cases it is not only the "resistance of the child" but the "resistance of the mother" that determines the extent of the influence of the other factors on the problem.

The influence of the character and intelligence of the mother on the rate of infant mortality is exceedingly difficult to measure statistically. It is especially difficult to distinguish this influence from that of the other closely related factors and one can never feel certain that the desideratum, "other things being equal," has been reasonably well attained. It is not to be expected, therefore, that the relationship between the character and intelligence of the mother and the rate of infant mortality can in the nature of the case be so accurately measured as to show the extent of the influence of the former on the latter, although the application of the statistical method does show the existence of such a relationship.

Thus, in the Boston investigation it was found that of the 341 mothers, upon whose general intelligence and standards of child care the investigators felt competent to express an opinion, the rate of mortality (99.6) was lower for the 254 infants of whose mothers a favorable opinion was given than the rate (103.4) for the 87 infants whose mothers were criticised unfavorably. These figures are of somewhat doubtful value, however, as they are based on an opinion formed upon only one visit and include only the somewhat extreme cases upon each end of the scale, all the other mothers in between being omitted, as the investigators did not feel competent to express an opinion upon them. A similar attempt

Digitized by Google

was made by Dr. Schwarz to grade the mothers coming under the observation of the New York Free Out-Door Maternity Clinic according to their general intelligence, which was more successful, as it was based on a much larger number of cases. Thus, among 2,326 infants whose mothers were rated as unsatisfactory in general intelligence the rate of mortality was 126 deaths per 1,000 births in comparison with a rate of 100 for the 210 infants whose mothers were rated as satisfactory in this respect.* These figures are also of doubtful value as the standard of intelligence used must have been high to have excluded such a large proportion of the mothers from the satisfactory grade.

A number of attempts have also been made to show the relationship of the intelligence of the parents to infant mortality in which literacy is used as a test of intelligence. Thus, Dr. Schwarz, in the study referred to above, shows that the rate of mortality among 1,297 infants, both of whose parents were literate, was 111 deaths per 1,000 births, while among 702 others, only one of whose parents was literate, the rate was 113, and among 458, where both were illiterate, it was 172.† In a similar manner it was shown in the Johnstown investigation that the rate of mortality was higher among infants born to foreign-born mothers who could not read and write any language or who could not speak English than among those infants whose mothers could meet these tests. Thus, of the 229 mothers who could not read and write, the infant mortality rate was 214, while among the 419 mothers who could read and write it was only 148. Among the 401 infants, whose mothers could not speak English, the mortality rate was 187 in comparison with a rate of 146 among those whose mothers could speak English.‡ In so far, therefore, as literacy and ability to speak English are indices of intelligence these figures indicate, though probably not adequately, the heavy handicap under which babies are placed because of a lack of intelligence on the part of their mothers.

* First Annual Report of the Free Out-Door Maternity Clinic, Covering the First Nine Years of the Clinic's Experience. New York, 1910; Chapter 2, Part 2, Report of the Pediatric Department, with a Study of Early Infant Mortality, by Herman Schwarz, M. D., p. 41.

† Loc. cit., p. 39.

‡ U. S. Children's Bureau: Infant Mortality: Johnstown, Pa. Washington, 1915, p. 34.

Digitized by Google

Other writers have shown the existence of this relationship between literacy and infant mortality by correlating the percentage of illiteracy and the rate of infant mortality in a number of cities. For instance, Phelps, in his recent study of the relation of infant mortality to the employment of mothers, concludes, after a careful examination of the available statistics for the cities and states of New England and especially of the cities of Massachusetts, that a high female illiteracy is "with fair uniformity coexistent with a high infant mortality rate." * Newman also notes the relationship from a study of English figures, although he does not find it to be as close as the writer just cited.†

On the whole, however, the extent of the influence of the character and intelligence of the mother upon infant mortality is not capable of statistical demonstration or of exact measurement by any method so far devised. Nevertheless observation and experience show that it is an all-important factor in the problem—the underlying factor which determines to a great extent the influence of other measurable factors. In conclusion, therefore, it may be said of this article, as Dr. Newman says of his own excellent work on the subject:

> This book will have been written in vain if it does not lay the emphasis of this problem upon the vital importance to the nation of its motherhood. Wherever we turn, and to whatever issue, in this question of infant mortality, we are faced with one all pervading need—the need of a higher standard of . . . motherhood. Infant mortality in the early weeks of life is evidently due in large measure to the physical condition of the mother, leading to prematurity and debility of the infant; and in the later months of the first year infant mortality appears to be due to unsatisfactory feeding of the infant. But from either point of view it becomes clear that the problem of infant mortality is not one of sanitation alone, or housing, or indeed of poverty as such, but *is mainly a question of motherhood.* No doubt external conditions as those named are influencing maternity, but they are, in the main, affecting the mother, and not the child. Improved sanitation, better housing, cheap and good food, domestic education, a

* Edward B. Phelps: Infant Mortality and Its Relation to Woman's Employment—A Study of Massachusetts Statistics. In Volume XIII, Part 1, of the Bureau of Labor's Report on Condition of Woman and Child Wage Earners in the United States, p. 48.

† George Newman, M. D.: Infant Mortality. London, 1906, p. 222.

Digitized by Google

healthy life of body and mind . . . exert but an indirect effect on the child itself, who depends for its life in the first twelve months, not upon the state or the municipality, nor yet upon this or that system of *creche* or milk-feeding, but upon the health, the intelligence, the devotion and maternal instinct of the mother. And if we would solve this great problem of infant mortality, it would appear that we must first obtain a higher standard of . . . motherhood. Without a moment's hesitation I place this need as the first requirement. Other things . . . are important, but this is the chief thing.*

To this conclusion the present writer heartily subscribes. The most important factors in infant mortality are the strength, the health, the character, and intelligence of the mother.

Preventive Methods: Education for Parenthood. Space would not permit, even if such a thing were possible, an enumeration of all the agencies and influences in modern life which are helping to raise the standard of motherhood or of all those other agencies and influences which are helping to lower it. It will be desirable, however, to conclude this discussion of the relation of the mother to infant mortality with at least a brief survey of some of the chief methods whereby the standard of motherhood, and also of fatherhood, may be raised. Probably the most important of these methods is education—the instruction and training of mothers and potential mothers in the care and rearing of children.

In a recent pamphlet, issued by the Children's Bureau on Baby-saving Campaigns, the methods employed by certain American cities in the prevention of infant mortality were pointed out—the work of health departments, registration of births and deaths, provision of a pure milk supply, milk stations and baby clinics, encouragement of maternal feeding, "little mother leagues" and classes, visiting nurses, prenatal work with expectant mothers, improvement of housing and living conditions, fight against flies, garbage accumulation, and dust, fresh air camps and hospitals, educational work through the distribution of circulars, pamphlets, etc.† A

* Ibid, pp. 257-258.

† U. S. Children's Bureau: Baby Saving Campaigns. Washington, 1911. See table of contents.

Digitized by Google

careful examination of each of these methods of preventing infant mortality will at once reveal to what a large extent each depends for its effectiveness upon the mother. At first sight it would appear that the chief function of a milk station is to provide cheap and pure milk and yet, as the report just referred to points out, "it has been the experience of practically all milk dispensaries that it is useless to send pure, clean milk into a dirty home to be handled by an ignorant, dirty mother or older child. It is necessary to reach the mothers, and not only teach them how to care for the baby's milk, but to convince them of the necessity of cleanliness. . . . In many cities it is believed that the principal good to be derived from milk stations consists in the opportunity given for those in charge to come in contact with the mothers and with the home surroundings. . . . Many cities place practically all the emphasis upon visiting nurses and the instruction of the mothers in the homes." * Dr. Josephine Baker, director of child hygiene of the Department of Health of the City of New York, also expressed the same idea in an address before the International Congress of Hygiene and Demography at Washington in 1912: "Without overlooking the value of pure milk, I believe this problem must primarily be solved by educational measures. In other words, the solution of the problem of infant mortality is 20 per cent. pure milk and 80 per cent. training of the mothers. The infants' milk stations will serve their wider usefulness when they become educational centers for prenatal instruction and the encouragement of breast feeding and teaching of better hygiene." †

In the case of visiting nursing, it is, of course, not necessary to point out that it is the function of such nurses not to care for the child themselves but to instruct the mother in the proper care of children.

Recently the work of visiting nurses has been extended to include the instruction of expectant mothers during the period of pregnancy. This "prenatal care," as it is often called, is

* Prev. cit., pp. 22–23, and 32.
† Transactions, Vol. 3, p. 149.

Digitized by Google

one of the newest and one of the most important direct means of reducing infant mortality by the education of the mothers, for the instruction is begun not after the child is born but soon after it has been conceived.

Finally, among the methods of instructing mothers in the proper care of children after the birth of the child or after conception, should be mentioned the educational work that is being carried on through the distribution of circulars, pamphlets, etc., on the care of infants and children as well as upon general hygiene and health. The volume of such publications, as well as the character of the material and the method of presentation and distribution, is improving rapidly. State and city boards of health and many private societies and associations have issued many exceedingly valuable pamphlets written in such a way as to appeal effectively to the mothers interested. Some of these have been printed in foreign languages as well as in English. The Federal Public Health Service and the Children's Bureau have also issued similar publications. The latter bureau is issuing a Care of Children Series, two numbers of which have already appeared. The first, " Prenatal Care," deals with the care of mothers during pregnancy, while the second, " Infant Care," gives the mother specific instruction as to the care of the baby after birth. Both of these pamphlets are written in popular, non-medical terms and are admirably suited to their purposes.

The trouble with the publications of the Children's Bureau, as well as those of the Public Health Service, is that the amount of funds available has heretofore been so small that they have not been as widely distributed as they should be. If these bulletins on the care of children, for instance, were as widely distributed as the Department of Agriculture's very successful series of Farmers' Bulletins, we should feel more confident that the government is doing its part in the prevention of infant mortality than is possible at present. On the other hand, the work of many city health departments has not been so limited by a lack of funds. In a number of cities educational pamphlets and folders are sent to each mother upon the registration of the birth of the child, either through the

Digitized by Google

mails or by the hands of a visiting or health nurse. Effective use has also been made of newspapers for the instruction of mothers in a number of cases.*

But this instruction of mothers in the care of themselves during pregnancy and of their babies after their birth is not the only means of reducing infant mortality by means of education. This kind of instruction may be called *instruction or education of parents;* but there is another kind of education that begins at childhood and which may be called *education for parenthood.* To say that "our educational system needs revision and that it should be made to conform more nearly to the actual requirements of our complex daily life" is merely to repeat a statement that has been made many times already; and yet there can be little doubt that one of the most effective means of lowering the rate of infant and child mortality, as well as the general death rate for all ages, would be the education of children and young men and women for parenthood. Just how much of this education should be given in the home and how much in the schools, and just how much should be obtained by the study of the laws of health and hygiene or by the study of domestic science and home management, it is impossible to say at present. What can be said at present is that the need of education for parenthood cannot be denied and that, in the overhauling and revision of our system of education which is going on at present, efficient methods for meeting this need must be worked out.†

But education for parenthood should include more than instruction in the science and art of infant and child care, or in home management or even in the ideals of parenthood; it should also include instruction, direct or indirect, in "the selection of parents." To take the case of the girl, for instance, education for parenthood,

* U. S. Children's Bureau: Baby-saving Campaigns, pp. 44–45 and Appendix.

† One of the most interesting of the recent attempts to instruct and educate children directly in the care of babies and in the laws and rules of infant hygiene has been carried on by the Department of Health of the City of New York in its lectures and classes for the "little mothers" who attend the city schools. Little Mothers' Leagues have also been organized which have reached a membership of over 17,000. These leagues were described by Dr. Baker in her address before the International Congress of Hygiene and Demography previously referred to. The idea has also been put in operation in other cities. Another movement of the same kind that should at least be mentioned is the establishment of continuation schools for homemaking.

Digitized by Google

instead of concerning itself with the care of her baby, will be at work when she is choosing the baby's father. In all times and places, woman's primal and supreme function is or should be that of choosing the father of the future. This great idea should be recognized, implicitly and explicitly, in the education of every girl; she is or may be partly responsible for the future of mankind. She herself, mind and body, is holy, for she is the temple of the life of this world to come. She must honor and care for herself accordingly; and this twofold aspect of her present and future duty, in caring for herself and in choosing her co-creator of the future, must be instilled into her mind with the solemnity, the sanctity, and the authoritative sanction of a religious dogma. *

This quotation, in suggesting how the standard of parenthood may be raised by eugenics, both on its positive side of the encouragement of the marriage of the fit and on its negative side of the discouragement or prevention of the marriage of the unfit, the unhealthy, and the weak, opens up an interesting field for discussion but one which, even if space would permit, it would hardly be proper for us to enter upon here.

* Saleeby, C. W.: The Methods of Race Regeneration. New York, 1911, p. 36.

Digitized by Google

MEASURES OF RELATIVE TAX BURDENS.*

By A. E. James, *Santa Fé, New Mexico.*

The larger problems of methods of taxation are receiving attention at a gathering of tax experts meeting in conference simultaneously with the gathering of this association. In their major deliberations the statistician has only a minor part. His duty is not so much to devise methods of reform as to measure the need of it—to point with the certain finger of fact to the weak spots in the present structure.

When the Committee on Tax Revision of the commonwealth of Virginia met to deliberate on the reforms it should recommend to the legislature whose creature it was, it felt imperatively the need for such guidance. The demand which had called such a body into being was primarily the product of disputes and recriminations between sections of the state, classes of its social and economic organization, and owners of its groups or kinds of property. Tidewater, Piedmont, Valley and Southwest, and the cities each claimed the others were tax dodgers and worse; the farmer and the merchant, the miner and the banker each was (in his own story and honestly-held opinion) the unjust bearer of his neighbor's burden; land and fluid capital, the widow's cow and the opulent railroad each groaned (so its owner said) under the intolerable discrimination of the tax system as administered by the officials of the commonwealth.

The committee held hearings. Each section, class, and group appeared, urged its grievance, and made its plea. But the taxpayer, whatever his honest convictions, usually found when forced to make his case that his facts were as few as his convictions were positive. The committee found that, if it listened to all pleas and took them at their face value, it would have to lower all taxes on everyone.

Here was the proper field for the statistician. All these claimants could not be right. The writer was employed to

* Paper presented at the meeting of the American Statistical Association, Stanford University, California, August 13, 1915.

Digitized by Google

endeavor in the short time at the disposal of the committee to find facts on whose general significance it could rely.

In undertaking this study the writer was confronted first with the limitation as to time. The entire work was to be completed between June 15 and November 1, 1914. The survey of the field must be rapid; it must also be reasonably accurate in its general results. It must cover the large items of taxable units in a large way and be content with that result. These units, as shown by the last preceding assessment, were as follows:

Item.	Assessed Value.	Per Cent. of Total.
Real estate	$550,421,094	56
Tangible farm property	44,382,756	5
Other tangible personal property	58,595,348	6
Banks	51,122,050	5
Railroads	109,310,445	11
Other public utilities	30,068,941	3
Intangible personal property	141,967,315	14
Total	$985,888,949	100

It is at once apparent that the first question relates to real estate. Is real property assessed at an equitable proportion as compared with other property, is it equitably assessed as between sections of the state, as between city and county, as between owners in the same locality?

At the outset it became necessary to assume two important propositions which a purely academic thesis could not assume: (1) that "equitably" meant equitably before the law and (2) that a tax meant any impost which must fall upon the property or the person because of the ownership of such property. The latter assumption was of prime importance in dealing with railroads and with certain other classes of taxpayers as we shall later see.

Two ways are possible to measure the ratio of realty assessments; a re-appraisal, and an examination of recorded sales or transfers. Obviously the former was impossible to us; the latter was the only choice. Conditions, moreover, made this a very easy method of practical application. A full sales report was on file with the state auditor; the detailed tax books were likewise on file in that office. The sales reports were first

Digitized by Google

edited to eliminate improper items; trades, transactions between relatives and nominal deals, sheriffs', executors', and guardians' sales, old deals lately recorded, deeds other than full warranties, etc. The comparison was restricted to a single year. After the first editing the sales remaining were copied on cards, then sorted and compared with the tax records for the entry of the assessed valuation. Thereupon a further editing was necessary to eliminate items not furnishing a comparable assessment. Thereupon the data were ready for tabulation.

The facts to be gleaned from tabulation were as follows—ratios of assessed to true values:

1. Between counties;
2. Between rural and urban property;
3. Between different classes of taxpayers in the same taxing district.

To bring out these facts the cards were sorted first into the one hundred counties and nineteen cities into which all Virginia is divided. Thereafter a division was made into platted and acreage property. For each of these grand divisions a composite tabulation form was made up classifying the sales first by white and colored ownership and second by the size of consideration paid for the property. The total assessments and considerations so tabulated and aggregated would make possible comparisons:

1. Between counties;
2. Between land and lots;
3. Between colored and white;
4. Between owners of large and small properties.

The results secured were interesting and conclusive. Speaking generally, the cities were assessed more nearly at true value than the country; Tidewater more than the interior; the Southwest mountain country was assessed the lowest of all. The inequalities ranged from a ratio of 12.5 per cent. in Carroll county to 76.5 per cent. in Fredericksburg City. Lots in counties were assessed on the average at 38.5 per cent. and lands in the same counties at 30.2 per cent.

Digitized by Google

As between large and small properties, the following table illustrates very well the inequality too often visited upon the small taxpayer:

TABLE I.

REAL ESTATE ASSESSMENTS AND SALES IN COUNTIES.

Value of Property.	Number of Sales.	Total Assessment.	Total Selling Price.	Ratio, Per Cent.
Total	16,362	$7,929,415	$23,666,730	33.5
Under $500	7,683	718,187	1,536,634	46.7
$500–$1,000	2,965	776,978	1,992,507	39.0
1,000– 2,500	3,219	1,791,904	4,924,387	36.4
2,500– 5,000	1,574	1,714,282	5,241,605	32.7
5,000–10,000	635	1,285,946	4,127,999	31.1
Over $10,000	286	1,642,118	5,843,498	28.1

REAL ESTATE ASSESSMENTS AND SALES IN CITIES.

Value of Property.	Number of Sales.	Total Assessment.	Total Selling Price.	Ratio, Per Cent.
Total	4,431	$6,475,430	$12,215,456	53.1
Under $500	921	143,974	241,443	59.8
$500–$1,000	835	335,129	575,904	58.2
1,000– 2,500	1,280	1,173,217	2,080,856	56.5
2,500– 5,000	856	1,555,122	2,776,501	56.0
5,000–10,000	361	1,263,713	2,383,685	53.0
Over $10,000	178	2,004,275	4,157,067	48.2

The foregoing table also points the way of reform. The inequality *among taxpayers* is far more vicious than that among *sections*, since the former applies to *all taxes*, the latter only to the relatively small state tax. Since, therefore, equality among taxpayers is the most necessary and since the higher assessment in cities was also accompanied by greater uniformity the first step in reform indicated by the statistics was a general increase in the ratio of assessed to true value.

The figures for colored and white owners were interesting, but for obvious reasons were not published by the committee. On their face they showed that colored owners were assessed more severely then their white neighbors. A more careful examination, however, revealed the fact that this result was due mainly to the smaller size of the properties owned by negroes. Within the same size groups the colored property was assessed somewhat higher than the white, but the average size of property within the groups was also less. I personally con-

Digitized by Google

cluded that the apparent discrimination against the Negro was due to economic rather than racial causes.

Turning from what we may call the internal comparison of real estate values to the external comparison, or, in other words, that of real with other property, we first considered tangible personal property.

No such easy measure and no such certain standard of value is available anywhere respecting personal property as is easily at hand in the case of real estate.

In the first place, while the entire omission of real estate is not unknown in states like Virginia, it is of infrequent occurrence. Not so with movable property; horses, cattle, sheep, swine, farm implements, merchants' stocks, manufacturers' stocks and equipment are all not merely subject to undervaluation, but to entire omission as well.

The United States Census was not used in dealing with real estate for the plain reason that better measures were available. In the case of personal property the census of farm property had to be used, compared with the assessed value of the same kinds of property, and from that basis an estimate made of the entire value of tangible property. As the same assessor valued all the items so estimated and would naturally try to do some measure of comparative justice it is possible that the error in these computations was within reason. The following table sets forth the final result of these estimates:

TABLE II.

VALUES OF TANGIBLE PERSONAL PROPERTY.

Classes of Property.	Numbers.		Values.		Average Values.		Ratio of Assessment.
	Assessment 1913.	Census 1910.	Assessment 1913.	Census 1910 and Estimated.	Assessment 1913.	Census 1910.	
Value of tangible property, all classes	.....	.....	$102,978,104	$220,300,380	.....	.....	46.80
Values of tangible property, classes not reported by census..	.....	.....	58,595,348	(estimated) 124,156,200 (Census of 1910)	.....	.....	47.20
Values of tangible property, classes reported by census ...	.....	.....	44,382,756	96,144,180	.....	.....	46.20
Horses, mules, etc..	393,273	433,837	26,784,959	48,083,529	$168.30	$110.90	55.80
Cattle.....	656,010	895,728	12,070,666	22,202,253	18.40	24.80	54.30
Sheep and goats...	339,218	787,595	1,011,858	3,340,087	2.99	4.24	33.30
Hogs.....	465,808	836,406	1,584,227	4,402,428	3.41	5.27	36.00
Farm machinery ..	.....	.....	2,931,36	18,115,883	.....	.....	16.19

Digitized by Google

The ratio of assessed to true value for real estate was 38.5 per cent., the ratio for *tangible* personal property was 46.8 per cent. Concerning this fact the report of the committee says:

The following conclusion can be drawn relative to the taxation of tangible personal property. This property on the whole is somewhat better assessed than is real estate, a fact not always clearly brought out in reports relative to taxation. Many writers have visited upon tangible personal property the odium very properly borne by the assessment and taxation of intangible personal property. As a matter of fact, tangible personal property partakes very largely of the condition of real estate. It is in view and most of it can be counted. It is available for appraisal in quantities whose values are fairly well known to the average man. The opportunities for under-assessment are rather less than those for the undervaluation of realty. The main defect of the assessment of tangible property arises rather from the failure to list the property at all than from the failure to list it at a reasonable value.

The same sectional inequalities found in real estate were likewise apparent in the ratios for tangible personal property. Warren county showed the lowest assessment, 23.3 per cent., Southhampton county the highest, 75.25 per cent. In round figures, horses averaged for the state $68 per head as assessed and $111 true value, cattle $18 and $24, sheep $3 and $4.25, hogs $3.40 and $5.26; farm machinery was assessed at $2,931,-036 and the reported true value was $18,115,883.

The estimate for tangible personal property, not enumerated in comparable form, was a somewhat doubtful procedure. Table II shows that the total assessed value of the comparable items was $44,382,756, and of the items not comparable with any census estimates, $58,595,348. The estimate of the true value of this large total was based upon the ratio in each county and city of the comparable items. One factor in favor of this procedure has been mentioned—the same assessor values all of the tangible schedule in Virginia.

Next to be considered and next in importance were the railroads.

The Virginia system of taxing railroads comprises:

1. A gross earnings tax of 1 per cent., the share of total

Digitized by Google

gross earnings accruing in the state being estimated on the basis of single track miles of main line.

2. The regular property tax imposed upon the real and personal property. The assessment is made by the corporation commission and apportioned among the counties, cities, and towns through which the roads operate.

To test the fairness of the above taxes it was necessary to measure:

1. The underassessment of railroads on a commercial basis.

2. The tax rate on full value taking account of the super gross earnings tax.

3. The average true value tax rate on other or comparable property.

Considering the last computation first it seemed best to make the comparison between railroads and real estate, for the reason that public service corporations should probably be equated with that class of property which forms the main basis of the revenue system. By dividing the tax imposed by all civil divisions upon real estate [(state $1,927,602.29, counties, cities and magisterial districts $6,006,352.93, towns $293,-537.38, total $8,227,492.60) by the true value already determined as described above (counties $976,698,395, cities $451,164,000, total $1,427,862, 395)] the real estate average tax rate was computed at 58 cents on the $100 of true value. By the extent to which the average for all railroads exceeded or fell short of this average it might be assumed that the roads were over- or under-taxed, ignoring for the moment questions of weighting.

The commercial values of the Virginia roads were computed on the stock and bond basis except where quotations were lacking. In such cases estimates were made upon the balance sheet and income account records as shown in the reports to the corporation commission. On the whole I personally lean toward the stock and bond method where the assets and income of the roads are based almost wholly (as is the case with the Virginia roads) on the operating property. The only danger is that the method will over-value the roads. Where the values so secured are to be contrasted with sales values of real estate which capitalize fully future prospects, I do not regard this as a serious objection.

Digitized by Google

The apportionment of the values computed for the system as a whole to the state of Virginia was made on the basis of net operating revenues as reported by the several roads for that state. While these reports were not made on a uniform basis it was not believed that this lack of uniformity produced any very considerable errors, particularly since the figure to be used was a proportion of Virginia net earnings to total net earnings, and only very substantial errors could materially affect the results.

The value estimated for the interstate roads was $357,333,700 and the taxes levied thereon were $2,227,026.15. This amounted to a rate of 62 cents on $100, or 4 cents in excess of the real estate rate. It was apparent, therefore, that the system was at least demanding sufficient revenue in total from the railroads. As between roads it was exceedingly unequal, the largest road paying 65 cents and three important properties only 38 cents, due in part to the arbitrary method of prorating gross earnings.

The study would have been more fully extended into individual localities, had time permitted, with a view to demonstrating local inequalities where the property locally assessed was appraised either higher or lower than the corporate prop-

TABLE III.

TAX STATISTICS OF ELECTRIC RAILWAYS AND LIGHT AND POWER OPERATIONS IN CONNECTION THEREWITH, YEAR ENDING JUNE 30, 1913.

Companies.	Taxes to $100 Stock and Bond Valuation.	Real Estate Tax on $100 True Value in Same Territory.
Appalachian Power Company (Interstate)	$2.26	$0.23
Blue Ridge Light & Power Co. (Staunton)	.93	.64
Charlottesville & Albemarle Rwy. Co.	1.33	.46
Danville Traction & Power Co.	.62	1.09
Lynchburg Traction & Light Co.	.72	.87
Mill Mountain Incline, Inc. (Roanoke)	1.06	.40
Newport News & Old Point Rwy. & Elec. Co.	.65	.69
Norfolk City & Suburban Rwy. Co.	3.92	.45
Radford Water Power Co.	1.22	.96
Richmond & Chesapeake Bay Rwy. Co.	2.71	.37
Richmond & Henrico Rwy. Co.	8.10	.93
Richmond & Rappahannock River Rwy. Co.	.54	.52
Roanoke Railway & Electric Co.	.58	.67
Tazewell Street Rwy. Co.	.52	.30
Virginia Railway & Power Co.	.76	.82
Total	$0.83	.....

Digitized by Google

erty assessed by the state corporation commission. This study was not made.

Minor public utilities were assessed, like the railroads, by the corporation commission and paid taxes to the local communities. The true test of their tax was not, of course, the general state average tax rate, but the true rate of the city or town in which they operated.

The danger of central assessment not closely associated with the local tax system was illustrated by the comparative rates for the principal companies. It should be noted that the Appalachian Power Company owned its main properties in Carroll county which county had the doubtful distinction of having the lowest assessment ratio in the state. As a result the valuation of the company for taxation was out of all decent proportion to the just value.

The analysis of bank taxation brought to light wholly undreamed of conditions of local assessments. The Virginia system calls for a report in duplicate—one to the state auditor and one to the local commissioner of the revenue—showing the book value of bank stock less real estate. The auditor collects the state tax directly from the banks; for the local tax they are placed on the local rolls. A special schedule sent to the banks showed taxes paid wholly at variance with the known factors—the book value and the local tax rate. It soon developed that a gentleman's agreement existed in many places under which the formal report was ignored and a more "equitable" one substituted. One county had even by resolution of the local board undertaken to exempt the banks therein from *all* local taxes—and the banks paid none.

Although the banks, despite these facts, were believed to be overtaxed, the committee felt a certain amount of such extra burden justifiable. The effort in this study was not to equate the bank tax with the tax on real estate, but to adopt a flat tax on all banks, granting a substantial but not a great reduction. The average tax on true investment was found to be $1.53 per $100; a reduction to $1.35 was recommended.

The shortness of the time forbade two investigations which I feel Virginia should certainly make. No careful study has

Digitized by Google

ever been made of the license system common to the whole South either from the economic or revenue aspects. The real burden of that system of taxing business is unknown either in terms of property or of income. An effort was made to secure such data, but the response was not satisfactory and the shortness of the time forbade further effort. The slight study made seemed to show that the system, while grossly evaded, was probably no worse in that respect than the property tax and was on the whole more certain of collection and less burdensome when once adjusted.

In one field no investigation was attempted at all. This was a study of the effect of the tax on intangible property. A glance at the table at the beginning of this paper will show that the assessment of this class of property is probably more nearly adequate in Virginia than in most states. This is due wholly to systematic record searching by a special appointive official—the examiner of the records. A comparison of appraisals of probated estates with the assessment books next preceding death would have been a valuable contribution to our knowledge as to the quantity of intangible property escaping taxation even where (judging by totals) it is most effectively reached.

The final study of general interest undertaken dealt with the taxation of insurance companies. In Virginia three overlapping taxes are imposed on insurance: (1) the tax on gross insurance premiums, common to most states; (2) an annual state corporation fee, heavy enough to be a material burden on some companies; and (3) municipal licenses in the form of an occupation tax on the business of the companies. Some of these municipal exactions are as high as $200 per company. Naturally since the licenses were not based upon property owned or business transacted they proved a serious burden upon small and new companies. Indeed the total burden was heavy upon all. This study was, therefore, directed to a comparison of premium income and taxes in Virginia. To equate insurance taxes with property taxes the admitted assets of the companies were apportioned to the state on the basis of premium income.

Digitized by Google

TABLE IV.

INSURANCE COMPANIES CLASSIFIED TO SHOW RELATIVE BURDEN OF VIRGINIA TAXES ON LARGE AND SMALL COMPANIES.

Classes on Basis of Premium Income.	Number.	Virginia Premium Income.	Virginia Taxes.	Percentage of Taxes to Income.
TOTAL OF ALL COMPANIES.				
Total all classes	173	$17,841,560	$440,784.74	2.47
Under $10,000	20	103,898	8,777.42	8.45
$10,000-$24,999	30	500,043	26,236.95	5.25
25,000- 49,999	48	1,744,481	75,135.32	4.31
50,000- 99,999	35	2,436,417	81,508.14	3.35
100,000-199,999	17	2,120,234	49,000.00	2.31
200,000-299,999	10	2,408,577	52,869.25	2.19
300,000-499,999	5	1,867,172	32,328.66	1.73
500,000 and over	8	6,660,738	114,928.91	1.72
FIRE AND MARINE.				
Total all classes	94	$4,984,824	$211,484.30	4.27
Under $10,000	8	48,756	4,687.41	9.61
$10,000-$24,999	20	3,311,914	19,889.21	6.42
25,000- 49,999	32	1,160,673	59,308.50	5.12
50,000- 99,999	23	1,500,231	63,175.63	4.21
100,000-199,999	6	787,133	25,274.96	3.21
200,000-299,999	5	1,176,117	39,148.59	3.33
LIFE.				
Total all classes	36	$9,026,399	$169,793.48	1.89
Under $10,000	2	3,345	856.42	25.62
$10,000-$24,999	4	81,359	2,365.15	2.91
25,000- 49,999	9	344,928	8,777.58	2.54
50,000- 99,999	4	324,958	6,785.68	2.04
100,000-199,999	5	615,576	11,652.58	1.89
200,000-299,999	2	511,023	9,530.74	1.87
300,000-499,999	3	1,203,384	23,701.81	1.97
500,000 and over	7	5,941,826	106,123.52	1.79
CASUALTY AND MISCELLANEOUS.				
Total all classes	32	$1,574,649	$40,257.84	2.56
Under $10,000	10	51,797	3,233.59	6.25
$10,000-$24,999	6	106,770	3,982.59	3.73
25,000- 49,999	6	210,661	6,467.37	3.07
150,000- 99,999	5	386,050	9,567.00	2.48
100,000-199,999	4	472,624	10,303.58	2.18
200,000-299,999	..	.....	.....	....
300,000-499,999	1	346,747	6,703.71	1.93
INDUSTRIAL AND SICK BENEFIT.				
Total all classes	11	$2,256,688	$19,249.12	.85
Under $10,0000	..	.....	.....	....
$10,000-$24,999	..	.....	.....	....
25,000- 49,999	1	28,219	581.87	2.06
50,000- 99,999	3	225.178	1,979.83	.88
100,000-199,999	2	244,901	1,768.97	.72
200,000-299,999	3	721,437	4,189.92	.58
300,000-499,999	1	317,041	1,923.14	.61
500,000-and over	1	718,912	8,805.39	1.23

Digitized by Google

TABLE V.

RELATIONSHIP TO PREMIUM INCOME AND TO ADMITTED ASSETS OF ALL VIRGINIA TAXES, FEES, AND CHARGES, EXCEPT ON REAL ESTATE, LEVIED ON INSURANCE COMPANIES.

Kinds of Insurance.	Premium Income.			Admitted Assets Less Real Estate.		Taxes.		
	Total.	In Virginia.	Per Cent. in Virginia.	Total.	Apportioned to Virginia on Basis of Premium Income.	Amount.	Per Cent. to Premium Income in Virginia.	Per Cent. to Admitted Assets in Virginia.
Fire and Marine Cos.	$253,065,001	$4,984,824	1.97	$488,866,059	$9,620,000	$211,484.30	4.27	221
Sick Benefit Cos.	3,943,634	2,255,688	57.25	932,712	533,000	19,249.12	.85	3.61
Life Cos.	599,494,978	9,026,399	1.51	4,018,923,606	60,550,000	169,793.48	1.88	0.27
Casualty, Fidelity and Miscellaneous	108,927,329	1,574,649	1.45	190,861,927	2,755,000	40,257.84	2.55	1.46
	$965,430,942	$17,841,560	1.85	$4,699,584,304	$73,458,000	$440,784.74	2.47	0.60

The regressive character of the tax on real estate appears again in insurance. The companies with premium receipts of less than $10,000 paid over 8 per cent. of their premiums in taxes, the largest less than 2 per cent. The Virginia taxes averaged about 2.5 per cent. of premium income. This charge was over 4 per cent. on fire and marine companies, less than 1 per cent. on sick benefit companies less than 2 per cent. on life companies, and a little over 2.5 per cent. on casualty and other insurance companies. Since premium income forms a less proportion of all income in life than in other forms of insurance the difference in favor of life companies was much greater than that shown here. This is most clearly illustrated by reducing the tax to terms of admitted assets. On this basis the average tax was 60 cents on $100, or about the same rate as railroads. It fell to 21 cents on life companies, however, and mounted to $2.21, or nearly ten times as much, on fire and marine.

To summarize: we found grossly unequal taxation in real estate as between large and small properties, as between city and country, from county to county, and between different sections of the state. We found the same inequalities in the assessments of tangible personal property. We found great variations among railroads and some difference between rail-

Digitized by Google

roads on the average and the average for real estate. We discovered gross mal-administration of bank taxes, unequal public utility assessments and illadjusted taxation of insurance companies, all by methods of investigation for which statistics are peculiarly fitted.

Taxation is peculiarly a field cursed by non-expert opinion and unenlightened legislative action. I know no field of statistical inquiry yielding more positive results or where honest investigation is more needed.

Digitized by Google

STATISTICS OF THE FOOD SUPPLY IN GERMANY.*

By Robert Morse Woodbury, Ph.D., *Cornell University.*

In the last fifty years Germany has changed from a food-exporting to a food-importing nation. Her agriculture, though protected in recent years by substantial tariff duties, has not been able to supply the growing demands for food-products arising from a largely increased population. Even before the war writers had pointed out the danger of a shortage of food in case of a European conflict. When the war came Germany found herself hemmed in on all sides by enemies, and cut off from her foreign sources of supply. Her foes have been unable as yet to make progress against the German armies; is it possible that they may achieve the same end, the defeat of Germany, by starving her out? Has Germany a sufficient home-production of food-stuffs to meet her most urgent needs?

The annual excess of imports over exports of food-products in normal times is considerable. In 1912 and 1913, years of exceptionally high home production, the excess of imports of grains and potatoes alone averaged 6,250,000 tons—2,000,000 tons of wheat, 3,000,000 tons of barley and 1,000,000 tons of corn, with smaller quantities of potatoes and rice. Meat and animal products, oil-cake and agricultural products used for fodder, were also imported in large quantities.†

Various estimates and guesses of the proportion of total consumption that is imported have been made. Low estimates of the dependence of Germany upon foreign supplies of food are based on the proportion, 10 per cent., of the bread-grains and 3–5 per cent. of the meat products, that is directly imported,‡ ignoring the fact that large amounts of meat and animal and vegetable products may be derived indirectly

* Paper read at the annual meeting of the American Statistical Association, Washington, D. C., December 29, 1915.

† Kuczynski and Zuntz, Deutschlands Nahrungs- und Futtermittel, Allgemeines Statistisches Archiv., Vol. IX, 1915, pp. 114, 137, 169, 173, 185–187.

‡ Ballod, Die Volksernaehrung in Krieg und Frieden, Schmollers Jahrbuch fuer Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reiche, Vol. 39, p. 77.

Digitized by Google

from abroad by import of fodder and fertilizers. Ballod, a German statistician who has written on the problem of the food supply, thinks that, including all kinds of vegetable food-stuffs imported, one third of all grains and vegetable food-products is imported; of meat products, one half comes directly or indirectly from abroad.* A careful and complete estimate has been made by Kuczynski and Zuntz, the former director of the Municipal Statistical Bureau of Berlin-Schöneberg, the latter director of the Institute of Animal Physiology of the Royal Agricultural High School in Berlin. In determining the proportion of food imported to the total consumed, the different kinds of food-stuffs are compared in terms of calories, or units of heat energy, contained in the respective foods; and also in terms of the amounts of protein, or nitrogeneous compounds, a certain quantity of which is an indispensable part of the daily ration. Of the total consumption of food in the years 1912–1913, these two authors estimate that 29 per cent. of the protein and 20 per cent. of the calories were imported. Of the total grains and plant foods used for human consumption, there was imported 9 per cent. of the heat energy and 15 per cent. of the protein content. Of the total meat supply 44 per cent. of the protein and 43 per cent. of the calories were imported either directly in the form of live cattle and dressed meats, or indirectly in the form of fodder for cattle and hogs.†

These figures, relating to food-stuffs used for human consumption, give some idea of Germany's present problem. Since the war, imports have been almost wholly cut off. Imports of wheat and barley from Russia and America have been stopped. Holland, Denmark, Switzerland, and Sweden have little to spare and Rumania has been unwilling to permit free export.‡ Austria-Hungary is in no position to help Germany out.§ In addition to the loss of foreign supplies, home production has been seriously affected by the war. Five hundred thousand horses have been taken from the farms for

* Ballod, Deutsche Volksernaehrung im Kriege, Preussische Jahrbuecher, Vol. 157, 1914, p. 102.

† Kuczynski and Zuntz, *op. cit.*, pp. 153, 131.

‡ Rumania forbade the export of rye, barley, and peas and levied a heavy export tax on the export of corn. Jahrbuecher fuer National-Oekonomie und Statistik, Vol. 105, July, 1915; Volkswirtschaftliche Kronik, pp. 293–4. See below, p. 29.

§ Kuczynski and Zuntz, *op. cit.*, p. 107.

Digitized by Google

the use of the army; and the supply can not be replenished from Russia, from which the greater part of the normal import, 114,000 horses, came.* Crops in small portions of Alsace and East Prussia were damaged by movements of hostile armies.† Farm labor has been called to the front.‡ About 18 per cent. of the total males and females engaged in agriculture are males between the ages 18 and 49 and therefore subject to military duty. The fisheries on the North Sea are practically closed down.§ The import of Chili saltpeter, representing one half of the total supply of nitrogeneous artificial fertilizer, has been cut off.|| The substitutes, a nitrogeneous compound manufactured as a by-product from coke, and nitrolin, made by electrical process from the atmosphere, are either insufficient in quantity or can be produced only at high expense.

Writers on the question of the food supply of Germany before the war were apt to be pessimistic. Since the war, in the many discussions of the food problem that have been published, the prevailing tone has been optimistic.¶ Though a large proportion of the food-stuffs consumed by the population are imported, yet in case of need there are substantial reserves to which the optimists point. Part of the abundant potato crop is now used for industrial purposes. There is annually a large surplus of sugar for export, and the sugar may be used to take the place of the missing imports, or the sugar lands may be used for other crops.** Finally, the grain used in the production of beer and starch can be used for food. In reply to these arguments it has been pointed out by Professor W. J. Ashley that the relative food values of these reserve

* Eltsbacher (ed.), Die deutsche Volksernaehrung und der englische Aushungerungsplan, 1914, p. 15.

† Gumbinnen and Allenstein in East Prussia were overrun by the Russians, and about 5–6 per cent. of the crop of Prussia was destroyed. Ballod, Die Volksernaehrung, Schmollers Jahrbuch, Vol. 39, p. 88. A small portion of Alsace-Lorraine was invaded by the French. Eltsbacher, *op. cit.*, p. 16.

‡ Eltsbacher suggests that "the greater part" of the farmers and farm laborers are in the "field." *Id.*, p. 14. In 1907, 3,146,592 males between the ages of 18 and 49 were engaged in agriculture, or 18 per cent. of the total males and females engaged in agriculture. Statistik des Deutschen Reiches, Vol. CCIII, pp. 2–3.

§ Eltsbacher, *op. cit.*, p. 14.

|| W. J. Ashley, Germany's Food Supply, The Quarterly Review, Oct., 1915, p. 446.

¶ See K. Borman, Die Lebensmittelversorgung des deutschen Volkes waehrend des Krieges, Deutsches Statistisches Zentralblatt, 1914, Vol. VI, column 286.

** Figures for sugar beet acreage for 1915 show a marked decrease, 31.9 per cent. over the acreage of 1914. Jahrbuecher f. Nat.-Oek. u. Stat., Vol. 105, Volkswirt. Kronik, p. 290.

Digitized by Google

crops must be taken into consideration.* Foods contain protein, carbohydrates, and fats in varying proportions. Protein includes the nitrogeneous compounds which are indispensable in rebuilding nervous tissue; neither of the other two constituents can take its place. There is a large proportion of protein in meats and the leguminous plants, and a fair proportion in wheat and other grains. Carbohydrates and fats are heat-producing. Potatoes are largely and sugar is almost exclusively composed of carbohydrates. The residues from the breweries, which contain a large proportion of the protein element of the grain used for the manufacture of beer, etc., have always been used as fodder for cattle, and thus indirectly have been utilized in the production of meat. It is therefore inadmissible to consider the total amount of grain used in the production of beer as a reserve.

An answer to the question, "Can Germany hold out?" has been attempted by a group of sixteen German statisticians, agriculturists, and physiologists and their conclusions have been published in a little volume edited by Eltzbacher,† the director of the School of Commerce of Berlin. Four main problems are investigated: (1) What was the total consumption of the population in 1912–13? (2) What is the total requirement of food of the German population? (3) What is the probable home-production of food-stuffs without considering any special emergency measures? (4) How much can the supply of food available be increased by special measures to be taken? The questions in which we are most especially interested are those relating to the statistical estimates of the food-supply. The statistical work seems to have been done largely if not entirely by Kuczynski and Zuntz.‡

The consumption for the total population of 1914 on the basis of the consumption, including normal imports, corresponding to the average of the years 1912 and 1913, would be 90,420,000,000,000 calories and 2,307,000 tons of protein. It is estimated that an adult man with moderate work daily re-

* W. J. Ashley, Germany's Food Supply, The Quarterly Review, Oct., 1915, pp. 444–462, esp. p. 447.

† Die deutsche Volksernaehrung und der englische Aushungerungsplan. Eine Denkschrift von Friedrich Aereboe, Karl Ballod, Frans Beyschlag, Wilhelm Caspari, Paul Eltzbacher, Hedwig Heyl, Paul Krusch, Robert Kuczynski, Kurt Lehmann, Otto Lemmermann, Karl Oppenheimer, Max Rubner, Kurt von Rumker, Bruno Tacke, Hermann Warmbold, und Nathan Zuntz. Edited by Paul Eltzbacher, 1914.

‡ See Kuczynski and Zuntz, *op. cit.*, note 2, p. 107.

Digitized by Google

quires food with heat energy of 3,000 calories, 11 per cent. of which, or 330 calories, is furnished by 80 grams of protein. Women and children require a somewhat less amount. Calculating the number of calories and the amount of protein required for the total population of 1914, the yearly physiological requirement for efficiency is estimated as 56,750,000,000,000 calories and 1,605,000 tons of protein. The consumption for the two years preceding the war exceeded the physiological requirement, calculated on this basis, by 59 per cent. for heat energy and 44 per cent. for protein.*

The probable home-production of food-stuffs is calculated from the figures of the production of 1912–13 by deducting 5 per cent. to allow for a decrease due to lack of labor, horses, fertilizer, to the cutting off of the fisheries, and to the damage to crops in East Prussia and Alsace. The total production is estimated at 1,554,000 tons of protein and 67,680,000,000,000 calories, a slight surplus of calories over the requirement but a deficit of 3 per cent. in protein.† The Eltzbacher report suggests a number of possible methods of making up this slight deficiency: lessening of household waste, slaughter of pigs and cattle to lessen the animal consumption of grains which might serve directly for human consumption, lessening of the use of grain in industry, development of the electric process for manufacturing nitrogen, culture of moorland, etc. By these measures it is calculated that 480,000 tons of protein can be added, transforming the deficit into a substantial surplus.‡

Professor W. J. Ashley, in a review of the Eltzbacher publication, criticizes these estimates on the ground that they are too optimistic. Wherever there is room for a difference of opinion as to the home-production or as to the requirement, he thinks that the authors have chosen the more favorable basis for estimate. His criticisms of the estimates are restricted mainly to the estimates of protein.

The figures for the supply of protein available from grains and potatoes are based on the assumption that the 1914 crop would be as good as the average of 1912 and 1913, with

* Eltzbacher, pp. 20–32, 64–66.
† Eltzbacher, *op. cit.*, p. 66–67.
‡ Id., p. 194.

Digitized by Google

an allowance of 5 per cent. for war conditions. Both of these crops, especially the 1913 harvest, were unusually good. The latest revision of the official estimates indicates, according to Ashley, that the 1914 crops fell about 13 per cent. short of the average of the two years preceding.* Furthermore, the official figures overestimate the crops. Ballod, a German statistician, thinks that the official figures are 10–15 per cent. in excess of the reality. Ashley reduces the figures for the protein content of grains and plant foods on the basis of a 15 per cent. reduction of the official figures for the 1914 crop from 963,000 to 753,000 tons. Estimates for the food value of cattle and hogs slaughtered are calculated by applying to the number of animals slaughtered the figures of the Health Department for the average "dead weight" per animal. But these averages, according to Ashley, are 10 per cent. too high. That they are too high is admitted in the Eltzbacher publication, but the averages are kept, "to include certain parts of the animal fats not fully included in the 'dead weights' though they are used for human consumption."† These animal fats are perhaps 5 per cent. of the total weight. Ashley reduces the estimate for protein 2 per cent. from 346,900 tons to 340,000 tons. He further criticizes the estimate for the annual average production of 2,200 liters (about 5,000 lbs.) of milk per cow and thinks that 2,000 or 1,800 liters is nearer the truth. He compromises, however, on 2,100 and reduces the estimate for protein in the milk supply by 5 per cent., or 8,300 tons. Accepting all the rest of the figures, these reductions would bring the estimate for the total available protein down from 1,554,000 tons to 1,411,000 tons.‡

Not satisfied with reducing the estimate of the supply available, Ashley revises the estimate for the physiological minimum necessary for the total population. The minimum requirement of protein necessary for efficiency as given by the "standard German author, Voit," is 118 grams of crude or 105 grams of digestible protein. Calculated on this basis the requirement would be not 1,605,000 but 2,001,000 tons of protein. Calculated on a basis of 92 grams of digestible

* Ashley, *op. cit.*, p. 449.

† Eltzbacher, p. 49.

‡ The deduction of 225,200 tons is made from 1,636,000 tons, the supply home-produced 1912–13.

Digitized by Google

protein, as suggested by Bauer* as the lowest figure consonant with efficiency, the requirement still would be 1,853,000 tons. To meet this requirement there would be available under normal conditions only 1,411,000 tons, a shortage of 24 per cent.

Finally Ashley believes that the estimate of 480,000 tons of protein added to the supply available by the measures of the government—the prohibition of the use of grain as fodder, reduction of the stock of pigs and cattle, agitation to decrease waste by housekeepers, etc.—is too high. He thinks the prohibition of the feeding of grain to stock was largely evaded at first on account of the large profits in meat production. There was a large percentage of waste incident to the storage of food by the government. A realization of one half of the saving hoped for would be, according to Professor Ashley, an outside estimate of the results achieved. On the other hand, Germany has been able to import or smuggle in food-stuffs by way of Holland and Denmark, the imports showing a very considerable increase over normal trade.

In conclusion Ashley calculates that there is a deficit of 12–13 per cent. of the food supply required—a deficit which is indicated by the great increase in the prices of food.† The abnormally high prices bring great suffering to the working classes in industrial districts. Meetings to discuss the prices of food-stuffs have been prohibited in Saxony. The agricultural population, on the other hand, has received some benefit from the high prices. Ashley thinks it was a mistake for the government to announce in an authoritative manner that there was food enough for all. If there is food enough, the question "Why should the prices of food-stuffs rise?" will naturally be asked among the working classes, with resentment against the inefficiency of the government in failing to stop speculation.

Are these criticisms of the German estimates well taken? If the German economists might tend to err on the side of overestimate, perhaps the Englishman might be inclined to be pessimistic about the food supply of the German people! No reference is made by Professor Ashley to an article which

* Handwoerterbuch der Staatswissenschaften, Vol. VI, p. 135.

† Seventy per cent. in Berlin in one year. Ashley, *op. cit.*, p. 455.

Digitized by Google

appeared in the *Allgemeines Statistisches Archiv* in the spring of the present year, written by Kuczynski and Zuntz, who were responsible for the statistical estimates of the Eltzbacher publication.* In the later article the estimates are much more detailed, critical comments, notes, and references to sources, which are omitted in the popular edition, are given, and in some cases, data which were not before available are used to revise the earlier figures.

The first question to be taken up in examining Ashley's criticisms in the light of the sources used by the German statisticians is the revision of the estimate of protein available from grains and plant-foods. Ashley reduced it from 963,000 to 753,000 tons, or 210,000 tons. Eltzbacher allowed for a 5 per cent. reduction from the production before the war to allow for poor crops—a deduction of 48,150,† so that the discrepancy between the two is some 161,800 tons. Kuczynski and Zuntz in their revised estimates increase the allowance for poor crops, lack of labor, horses and fertilizers, losses due to the movements of hostile armies, etc., to 8 per cent. This reduces their earlier figures by 28,900 tons. The actual returns of the 1914 crop show that even this increased allowance was too low; actually the 1914 crop fell some 12 per cent. below the figures for the average of 1912 and 1913.‡ A further reduction should be made, therefore, in Kuczynski and Zuntz's figures of 30,000 tons of protein.

The remaining difference of approximately 100,000 tons is due to Ashley's reduction of the crop estimate for 1914 by 15 per cent. to allow for error in reporting the yield. In support of this procedure he appeals to the criticism of crop reports by Ballod.§ Estimates of crops in Germany, in normal times, are returned in November by a large number of agricultural experts, each of whom estimates or guesses at the average yield per acre of the principal crops in his district. The districts now average nearly 16 square miles. The figures of average yield are multiplied by the area under cultivation in each crop as reported by the communal authorities.

* See above, p. 96, note 3.

†The deduction of 5 per cent. was not restricted solely to grains, etc., used for human consumption, but included also 33,850 tons for indirect use of grains and fodder.

‡A reduction of 12 for wheat; 12.5 for rye; 12.3 for barley, 0.8 for oats, and 14 per cent. for potatoes. Ashley's 13 per cent. reduction is therefore only slightly overstated.

§ Ashley, *op. cit.*, p. 449, note. See also the two articles by Ballod already cited.

Digitized by Google

Estimates of acreage can be compared with the results of agricultural censuses—in 1907 in Bavaria the estimates were slightly higher than the census figures.* The estimates of the communal authorities for the acreage in 1913, on the other hand, were almost uniformly lower than the results of the agricultural census of 1913.†

With reference to the estimates of average yield per acre, there is no direct way to test the figures. No agricultural census of crop production has ever been taken in Germany. In the United States estimates of crops are compared with census returns every 10 years and usually the census figures are taken as a new base for future estimates; each year the crop reporters report not an average yield for the district but the percentage yield of the present in terms of last year's crop. In Germany, there was no base to start from, the estimates have never been compared with census figures, and no one knows how far the estimated yields per hectar depart from the truth.

From 1893 to 1898 there were two different sets of crop estimates, one made by the agricultural experts (landwirtschaftliche Sachverständige), as at present, and the other by communal authorities, in most states usually in February after the harvest. In 1899 the second method was dropped, because the figures so obtained were almost out of date before they were published. These two estimates differed regularly by a considerable margin. The figures of the agricultural experts were usually 12–19 per cent. higher than the other figures.‡ It is suggested on the one hand that the experts,

* Huber, Die uebrige landwirtschaftliche Statistik, in Die Statistik in Deutschland nach ihrem heutigen Stand. Ed by Zahn, Vol. II, p. 140.

† The difference is small, not over 1 or 2 per cent. For the results of the Bodenbenutzungsaufnahme see Jahrbuecher fuer Nat. Oek. und Stat., Vol. 105. Volkswirtschaftliche Kronik, p. 295 ff. The estimates are found in the Vierteljahrsheft zur Statistik des deutschen Reichs., Vol. 23, 1914, I, p. 130.

‡ See Ballod, Gueterbedarf und Konsumtion in Die Statistik in Deutschland, Vol. II, p. 610. See also Morgenroth, Die Lebensmittelversorgung des deutschen Volkes waehrend des Krieges, Deutsches Statistisches Zentralblatt, Vol. 6, 1914, column 323. The discrepancy between the two sets of figures is even greater than Ballod states it. In Bavaria and Wuerttemberg and some other small states in this period both sets of crop estimates were made by agricultural experts. The difference in the two estimates in Prussia in 1894 varied from 13 per cent. for winter wheat to 33 per cent. for oats. By averaging in Bavaria and Wuerttemberg where the difference in the estimates was very small, the average difference is considerably reduced. The reason for the difference in these states was the slightly different time at which the two estimates were made. Since 1899, the number of crop reporters has been nearly doubled, a circumstance that might lead to greater accuracy. See Vierteljahrshefte zur Statistik des Deutschen Reiches, 1894, Vol. III, part III, pp. 38–48; part IV, p. 140; 1899, Vol. VIII, part III, p. 65; Ergaenzungsheft to part IV, p. 9, 16.

Digitized by Google

where farmers, would tend to generalize from their own good crops, and on the other that the returns gathered by the communal authorities would show understatement due to fear of taxation on the basis of the returns.

Which of these estimates is nearer the truth? How far do the estimates of the agricultural experts overstate the real harvest? The only piece of convincing evidence on the amount of overstatement is contained in an article by Ballod, which appeared in Schmoller's Jahrbuch for the present year and to which Ashley does not refer.* The amount of wheat and rye ground in mills plus the amount used for seed plus the amount fed to live-stock should equal the total amount available—the harvest plus the excess of imports over exports. The amount fed to cattle must be estimated roughly. Ballod thinks that 2,000,000–2,500,000 tons is an outside estimate of the rye used for fodder. He finds, however, that there is an excess of rye used either for fodder or due to overestimate of crops of 3,240,000 tons, or 30 per cent. of the harvest for 1908–1909, and of 3,840,000 tons, or 34 per cent., for 1909–1910. A similar excess of wheat, 620,000 tons and 1,170,000 tons, respectively, or 15 and 28 per cent. of the harvests, is found. Ballod explains the discrepancy as due to an overestimate of the rye and wheat crops of 10–15 per cent. Even making some allowance for understatement in the statistics of grist-mill production, the overestimate would still be considerable.

Ashley reduced the Eltzbacher estimates of protein from grains by 15 per cent., the larger of the two figures previously given by Ballod. Eltzbacher, however, makes an even more generous allowance than Ballod for wheat and rye used as fodder and in industry before calculating the amount of grain available for human consumption. For rye he deducts 2,708,-085 tons and for wheat 417,384 tons, or 23 and 9 per cent. of the harvest, respectively. Assuming that the percentages of the crop rather than the amount used as fodder are correct,

* In the article to which Ashley refers especially, the harvest of wheat and rye minus exports in four small East Prussian districts was expressed as a per capita consumption of bread. The consumption appeared to Ballod to be far too high. No allowance was made for use of wheat or rye as fodder or for industrial purposes. Further, an overstatement of yield in one district would not be at all conclusive as to a general overestimate. See Ballod, Deutsche Volksernaehrung im Kriege. Preussische Jahrbuecher, Vol. 157, 1914, pp. 103ff.

Digitized by Google

the basis of Ballod's calculation of overestimate is completely altered and without making any allowance for understatement in the statistics of grist-mill production the maximum error in the crop estimates would be 9 per cent. for rye and 11 per cent. for wheat, and not 15 per cent. as Ashley assumes.

The two authors, in the revision of the estimates, increase the allowance for rye used as fodder from 25 per cent. to 30 per cent. of the net harvest (after seed is deducted). This change means a reduction of 25,400 tons of protein in the supply available for human consumption. The possible overestimate of protein available from the rye crop is thus reduced to 4 per cent. There is no justification for a reduction in the crop estimates in view of this change, of more than 7.5 per cent. and a reduction of as much as that would hardly seem to be warranted on the basis of the figures given.*

On the other hand, Kuczynski and Zuntz make two important changes in their estimates, both in the direction of reducing their earlier figures. The total production of buckwheat, millet, and maize is revised, so that the protein available from these crops is reduced by 6,800 tons. The estimate of the production of beans, peas, and lentils is reduced almost half, involving a reduction in protein of 33,800 tons, or a total reduction of 40,000 tons.

The other differences in the estimates of protein available are of minor importance and are cited only to illustrate the statement that all the calculations were too optimistic. The two authors cite, as the basis of their estimate of the annual milk production per cow, the official estimates of Bavaria (1907) of 2,340 liters and of Württemberg of 1,940 liters.† Bavaria has nearly three times as many cows as Württemberg; by weighting the figures in proportion to the number of cows in each state an average of some 2,240 liters is secured. But the higher of these two figures, and that to which the greatest weight is attached, is an average of guesses made for the dif-

* The protein from wheat and rye accounts for about three fourths of the protein derived directly from grains and potatoes. The question can be waived as to whether the estimates for the proportion of the crop used as fodder are too high. If they are, they absorb, so to speak, an error in the crop estimates. The net result is to lower the amount of protein calculated as available for human consumption just as effectively as by a reduction in estimated harvest. No change has been made in the supply of meat available as a result of increasing the estimate of the amount of rye used as fodder.

† Kuczynski and Zuntz, *op. cit.*, p. 144.

Digitized by Google

ferent provinces by agricultural experts in consultation with inspectors of breeding, an altogether unsafe basis for a generalization and one that would probably err on the side of overestimation.*

With reference to the estimates of "dead weight" used to figure the total meat supply, it is admitted by the two authors that the figures used are too high, but the use is justified on the ground that some animal fats used for human consumption are not included in the "dead weights." A note states that the usual estimates of protein content are too high and of fats too low. It seems inferable that in expressing the food value in protein and calories a correction is made to give the true composition of the "dead weights."†

Ashley's criticism based on the overestimate of the "dead weights" loses its force if such a correction is made.

To sum up this discussion of the estimates of protein available: the final results do not differ much from Ashley's conclusion. Of the 162,000 tons discrepancy in the protein from grains and potatoes, an excess of 94,000 tons is admitted by the two authors in their revision and of 30,000 tons more is proved by the actual figures of the 1914 crops. Disregarding the minor items of protein in milk and the overestimate involved in the "dead weights," if a further reduction of 20,000 to 50,000 tons is made for possible crop overestimates the amount of protein available, 1,444,000 to 1,414,000 tons, corresponds fairly closely with Ashley's results. This computation leaves out of account a further reduction of protein available due to

* Die Milchwirtschaft in Bayern, Heft 78 der Beitraege zur Statistik des Koenigreichs Bayern. Hrsg. v. K. Statistischen Landesamt, pp. 18–19.

A reference is given also to a later estimate for Wuerttemberg (1910) of 2,128 liters. This figure is based on the result of milk tests on 344 cows of three breeds on 59 farms in different parts of the kingdom. The number of tests is altogether too small to form a reliable basis for generalization. Milk tests are more likely to be carried out on farms where production is above the average. Tests, for example, of 834 cows in Jefferson County, N. Y., showed an average annual production of 6,621 lbs. as compared with census figures for the county for 1909 of only 4,310 lbs. Milk tests in Bavaria are carried out on cows used for breeding.

See Die Milchversorgung in Wuerttemberg, O. Truedinger, 1914, Schriften des Vereins fuer Sozialpolitik, Vol. 140, Part IV, pp. 19–20. Truedinger made the earlier estimate for Wuerttemberg of 1,940 liters. See also H. H. Wing, The Formation of Cow-Testing Associations. Circular No. 17, April, 1913, Cornell University Agricultural Experiment Station of the College of Agriculture, Dept. of Animal Husbandry, p. 54. Figures for the tests in Jefferson County were kindly furnished me by Professor G. F. Warren, Cornell University.

† Kuczynski and Zuntz, *op. cit.*, p. 135, note, p. 139. "Infolgedessen wird meistens ein zu hoher Eiweissgehalt und ein viel zu geringer Fettgehalt angegeben. . . . Hier wurde tunlichst die durchschnittliche Zusammensetzung im 'Schlachtgewicht' berueeksichtigt."

Digitized by Google

possible loss under war conditions of home-grown fodder materials, for which Eltzbacher made an allowance of some 38,000 tons of protein.

The two authors practically concede in advance the justice of Ashley's criticism of the estimates of daily requirement of protein. After giving 70 to 80 grams as the physiological minimum for efficiency, 10 per cent. is added to allow for unavoidable waste and 5 per cent. more for a certain "excess-nourishment," in all approximately 92 grams. Ashley accepts 92 grams as the lowest reasonable minimum.

The final results of the two authors are much less optimistic than the conclusions of the Eltzbacher publication. They find a deficit of home-production of 21 per cent. of the requirement of protein and 10 per cent. of the requirement of calories; and this requirement itself means 16 per cent. less of protein and 19 per cent. less of calories than was consumed on the average before the war.

There remains the question as to the possible additions to the supply of protein which can be expected from the special measures of the government. Most of the estimates are pure guesses, without any trustworthy statistical basis. Almost half, 217,000 tons, of the total addition (480,000 tons) expected is to be achieved by changes in the keeping of livestock, lessening the waste of fodder and reducing the daily ration of some 3,000,000 "overfed" cows. The most dubious assumption in the calculation is that a saving of 249,000 tons of crude protein can be affected in the fodder ration without lessening at all the average production of milk of 2,200 liters. A calculation is made that the stock of pigs will have to be reduced by one third and of cows by one tenth as a consequence of the cutting off of the import of foreign fodder materials.* It is estimated that the prohibition of the use of bread-grains as fodder will add, if observed, 78,300 tons of protein to the supply available. Ashley thinks, and probably correctly, that the prohibition was largely evaded at first. Furthermore to reckon the entire amount of protein available in the grain estimated as used for fodder reopens the question as to a possible overestimate of this amount and of an overestimate of crop returns. In de-

* Eltzbacher, *op. cit.*, pp. 121–122. The elimination of the poorest tenth of the cows might tend to raise the average production.

Digitized by Google

fault of more certain evidence on these points, a somewhat liberal margin must be allowed.

In the revision, the two authors, if they were responsible for the estimates, seem to have realized the doubtful value of some of them. The attempt to calculate the amount added by the different measures of the government is abandoned completely. Attention is centered on the question of the supply of fodder materials, but the method of calculation is wholly altered. Instead of estimating the number of pigs and cattle that were dependent on the supplies of foreign fodder, a laborious estimate is made of the probable supply of the fodder materials of the Empire and of the total requirement of each kind of live-stock. They find that the total heat energy of fodder consumed by live-stock in 1912–13 was double the heat energy of foods used for human consumption. The actual consumption of fodder in these two years exceeded the minimum requirement as calculated by the two authors by 14 per cent. of crude protein and 8 per cent. of calories. One half of this difference might be saved by avoiding waste and reducing over-feeding: the remainder, the unavoidable waste, is added to the minimum requirement to make the practical requirement. During the war, with imports of fodder cut off, the supply is 13 per cent. below the practical requirement of crude protein and 12 per cent. below that of calories. If the entire deficit of calories and protein for both men and animals is reduced to a deficit of fodder, eliminating double counting, the available supply of protein is found to be 20 per cent. and that of calories 18 per cent. less than the practical requirement. The conclusion is drawn that the only way out of the difficulty is to reduce still further the number of cattle and pigs.

In summary the criticisms of Ashley seem to have been anticipated by the two authors in their revision, and it seems to me that Ashley's conclusions, though based in part on unsatisfactory or insufficient evidence, are substantially correct.

A few words should be said about the limitations of these statistics. They apply only to Germany proper. No account is taken of the areas of hostile territory occupied by the German armies, Belgium, Northern France, and Russian

Digitized by Google

Poland, nor of the populations of these districts. No deduction from the total food requirement is made for the possibility that the German armies operating in enemy territory could live off the country, and no addition to the requirement of food is made for prisoners of war held in Germany. The population is assumed constant from July, 1914. No account is taken further of the supplies of grain on hand at the end of the season left over from the 1913 crop. It may be that in these omissions some important factor has been disregarded.

It may be suggested that the amount of protein physiologically necessary is a quantity which has been variously estimated and which therefore offers no certain standard by which to test a deficit of food supply. Estimates are usually based either on statistics of consumption of healthy and efficient laborers or on laboratory experiments. A very low estimate of 50–60 grams daily was made by Professor Chittenden of Yale.* He made some experiments in which he showed that the norm previously accepted as the minimum requirement of protein was much higher than was absolutely necessary. It may be noted in passing that though Ashley and Kuczynski and Zuntz agree on a practical requirement of 92 grams of protein, the two authors include in it an allowance of 10 per cent. for unavoidable waste. A test to which no objection could be made would be to compare the available supply with the consumption of protein before the war of approximately 110 grams.†

In a computation of the available food supply, to deal with averages is not an entirely correct procedure. The food supply is not divided up among the population equally. Even in normal times there are some who do not get enough to maintain efficiency, and there is a large class who can afford no surplus of nourishment. For these classes a shortage of food supply means immediate loss of efficiency or suffering or both, and for these classes an increase of prices is tantamount, for

*Professor Chittenden's experiments are mentioned by Eltsbacher in defense of his own low figure of 80 grams.

†110 grams for an adult male with moderate work. Per capita of the actual population 83.4 grams were consumed before the war, as compared with 66.7 grams reckoned on a basis of about 92 grams per adult male. See Kuczynski and Zuntz, *op. cit.*, pp. 156–158.

Digitized by Google

all practical purposes, to a shortage of food. In spite of the maximum price regulations adopted by the German government, prices of food-stuffs in Berlin in one year of war have risen 70 per cent. since July, 1914.* The well-to-do classes, do not have to limit seriously their consumption of food except in case of those food-stuffs, notably bread and milk, the consumption of which is regulated by the card system with a fixed maximum amount which any one person may buy.

But the situation has been materially changed in some respects since these estimates were made. Forecasts of the new crops indicate a much better harvest, especially of wheat and rye, than was secured in 1914. The situation in Austria-Hungary will be much improved by the success of the wheat crop of Hungary, the forecasts of which indicate a very considerable increase over that of last year. Austria normally imports some 300,000 tons of wheat: the relative failure of the 1914 harvest in Hungary made the situation much worse than usual. Germany's harvests are much more like an average crop, though probably not equal to the yields of 1912 and 1913. In another year of war, Germany's deficit will therefore be smaller than that calculated for 1914.

Furthermore, Germany has been able to import considerable supplies of animal food products from Holland and Denmark. Figures, which Ashley gives, indicate a somewhat larger import than in normal times of butter, eggs, cheese, and fish.† Germany takes almost the whole of the surplus of the Dutch fisheries.

The success of the operations against Servia has opened up Bulgaria's reserves of wheat and other grains. Her normal exports of wheat amount to nearly 300,000 tons, or 7 per cent. of Germany's home production. Some of her surplus normally goes to Austria. If all were sent to Germany and all used for food, it alone would mean an addition of 20,300 tons to the supply of protein. In addition there is a considerable surplus of corn and barley and a surplus available for immediate needs of a somewhat uncertain amount left in the country from last year's crops. Bulgaria exports an average

* Ashley, *op. cit.*, p. 455.

† Ashley, *op. cit.*, p. 454. See also New York Times, Dec. 17, 1915.

Digitized by Google

of $5,000,000 worth of animal products yearly. Negotiations with Rumania for the sale of large quantities of grain have been completed.* Rumania's surplus of wheat is much larger than that of Bulgaria, amounting normally to 1,460,000 tons.† This would mean an addition to the annual supply of nearly 100,000 tons of protein. Besides, there are large quantities of corn and barley which can be used to supply the deficiency in fodder. A rough estimate would place the surplus of these crops available for export at one fourth to one third of the deficit in Germany. With these new sources of fodder materials opened, it is possible that a larger number of pigs and cattle were slaughtered than was necessary.

In conclusion, there is little possibility that Germany can be starved. Her supplies are probably sufficient to cover the minimum practical requirement. Her food supply falls considerably short of the actual rate of consumption in 1912–13, and it must be conserved and carefully distributed to ensure a sufficiency in the last months before the new harvests are available. There are distinctly less meat and animal products available than in normal times. The serious danger, it seems to me, is that disaffection may be spread among the working classes by restrictions on the consumption of food-stuffs, and above all by the increase in prices, which may seem to them entirely unnecessary if, as the government has announced, there is really enough food for all.

* See New York Times, news items in issues of Dec. 23 and Dec. 28, 1915.

† Rumania's crop for 1915 is estimated to be 24 per cent. larger than the average of 1909–13. These figures for the normal exports are taken from the Monthly Crop Report, published by authority of the Secretary of Agriculture, Vol. I, No. 5, p. 48, Sept. 15, 1915. They are based on the average of the last 5 years, 1909–13, as reported by the International Institute of Agriculture at Rome. In calculating the available protein, the same process has been followed as is used by Kuczynski and Zuntz.

Digitized by Google

PROCEEDINGS OF THE SEVENTY-SEVENTH ANNUAL MEETING OF THE AMERICAN STATISTICAL ASSOCIATION, WASHINGTON, D. C., DECEMBER 28-30, 1915.

The seventy-seventh annual meeting of the American Statistical Association was held at Washington, D. C., December 28-30, 1915. The following program was carried out:

PROGRAM.

TUESDAY, DECEMBER 28.

8.00 p. m. First Session, Large Banquet Hall, Hotel Raleigh.

Joint Meeting with the American Sociological Society.

Presidential addresses:

E. Dana Durand, President of the American Statistical Association, "The Emergence of Unearned Increment."

Edward A. Ross, President of the American Sociological Society, "War as Determiner."

WEDNESDAY, DECEMBER 29.

10.00 a. m. Second Session, Hotel Shoreham.

Joint Meeting with the American Political Science Association; general topic, *Standardization and Government Efficiency;* presiding officer, Morris Llewellyn Cooke, Director, Department of Public Works, Philadelphia.

1. Standardization in Statistical Presentation, Roland P. Falkner, Alexander Hamilton Institute.

Discussion: Leonard P. Ayres, Russell Sage Foundation, Julius H. Parmelee, Bureau of Railway Economics, Washington, Charles H. Verrill, U. S. Bureau of Labor Statistics.

2. The Meaning of Standardization, J. Russell Smith, University of Pennsylvania.
3. Standardization of Grades and Salaries in Civil Service, Robert S. Moses, Bureau of Municipal Research, New York.

Discussion: Henry Bruère, Chamberlain, New York City, J. A. Dunaway, University of Pennsylvania.

2.30 p. m. Third Session, Small Banquet Room, Hotel Raleigh.

1. Statistics of Imports and Exports, Frank Roy Rutter, Bureau of Foreign and Domestic Commerce.

Discussion: N. I. Stone, Brooklyn, N. Y., Henry C. Emery, New Haven, Conn., Martin J. Shugrue, Boston, Mass.

Digitized by Google

2. Standards for Graphic Forms of Presentation. Discussion of the Preliminary Report of the Joint Committee on this subject representing various national scientific associations including the American Statistical Association.

Discussion to be opened by Leonard P. Ayres, Russell Sage Foundation, Secretary of the Committee.

3. Agricultural Element in the Population, Eugene Merritt, U. S. Department of Agriculture.
4. Statistics of the Food Supply in Germany, Robert M. Woodbury, Cornell University.
5. Longevity of Presidents and Members of Congress, Irving Fisher, Yale University.
6. Business Session.

THURSDAY, DECEMBER 30.

Joint Meeting, by invitation, with Subsection B, Section VIII, of the Pan American Scientific Congress.

10.00 a. m. Fourth Session, Oak Room, Hotel Raleigh.

1. The Nature and Significance of the Changes in the Birth and Death Rates in Recent Years, Walter F. Willcox, Cornell University.
2. The Influence of Vital Statistics upon the Conservation of Human Life, W. S. Rankin, Secretary of State Board of Health, North Carolina.
3. The Relation of Sickness Reports to Health Administration, John W. Trask, Assistant Surgeon General, U. S. Public Health Service.
4. Vital Statistics in Relation to Life Insurance, Louis I. Dublin, Statistician, Metropolitan Life Insurance Company.
5. Statistics of Infant Mortality, Lewis Meriam, Bureau of Municipal Research, New York City.

2.30 p. m. Fifth Session, Oak Room, Hotel Raleigh.

1. The Federal Registration Service of the United States: Its Development, Problems, and Defects, Cressy L. Wilbur, Director, Division of Vital Statistics, New York State Department of Health.
2. Vital Statistics in the States and Cities of the United States, Wilmer R. Batt, State Registrar of Vital Statistics, Pennsylvania, Charles V. Chapin, Superintendent of Public Health, Rhode Island, William H. Guilfoy, Registrar of Records, New York City.
3. Informe sobre el desarrollo de la estadistica demografica de El Salvador, Dr. Pedro S. Fonseca, of Salvador.
4. The Accuracy and Completeness of Compiled Vital Statistics in the United States, John S. Fulton, Secretary of State Board of Health, Maryland.

Digitized by Google

Minutes of the Business Meeting.

Meeting called to order by President Durand at 6 p. m.

The reading of the minutes of the previous annual meeting was dispensed with as they had already been published in the proceedings.

The Secretary presented the following report:

Secretary's Report.

Mr. President and Members of the Association:

I have the following to report in regard to the condition of the Association:

Membership December 30, 1914, date of last annual meeting		681
New members added		75
Deaths during the year	12	
Resignations	22	
Dropped	39	
Total deductions		73
Present membership		683
Net gain		2
Beside members we have:		
Subscribers (mostly libraries)		205
Domestic Exchanges		38
Foreign Exchanges		117
Total mailing list		1,043

The death of the following members has been reported during the year: Honorary Member; Professor Ivan Iaanschul; Fellow, Edward Bunnell Phelps; Members, Nelson W. Aldrich, Walter F. Allen, A. C. Dale, C. R. Henderson, Oscar B. Ireland, Thomas Lawless, Arthur T. Lyman, George E. Millichamp, Peter White, Stuart Wood.

Four regular numbers of the Quarterly Publications have been issued during the year, containing an aggregate of 481 pages, or an average of 120 pages, per number. Volume XIV, which is completed with the December number of the

Digitized by Google

present year, contains 831 pages, or an average of 104 pages per number. This is the largest volume the Association has ever issued.

Four meetings of the Association have been held during the past year; one of these was the special meeting held in San Francisco in connection with the Panama Pacific International Exposition, August 11–13. The other three meetings were what we formerly called quarterly meetings; two of these were held in New York City, on April 15 and December 3, respectively; the third one was held in Washington, D. C., April 17. These meetings took the form of a dinner, followed by round-table discussion. The average attendance was about 40.

Respectfully submitted,
CARROLL W. DOTEN,
Secretary.

On motion, it was voted to accept the report of the Secretary and to authorize its publication in the proceedings in the March number of the QUARTERLY PUBLICATIONS.

TREASURER'S REPORT.

December 22, 1914, to December 23, 1915.

RECEIPTS.		
Membership dues	$1,242.32	
Sales and subscriptions	1,369.28	
Dividends and interest	143.73	
Balance on hand, December 22, 1914	741.30	
		$3,496.63
EXPENDITURES.		
Printing	$2,111.33	
Postage	129.45	
Salaries and clerical service	401.28	
Expenses	13.10	
	$2,655.16	
Balance on hand, December 23, 1915	841.47	
		$3,496.63

Digitized by Google

ASSETS.

17 shares B. & A. R. R. stock @ 189 per share, $3,213.00.

S. B. PEARMAIN,
Treasurer.

AUDITORS' REPORT.

We have audited the accounts of the treasurer of the American Statistical Association, for the year ending December 23, 1915, and counted the securities in his possession. We find the accounts accurately stated and the expenditures properly vouched.

Respectfully submitted,
LEROY D. PEAVEY,
MARTIN J. SHUGRUE,
Auditing Committee.

SEVENTY-FIFTH ANNIVERSARY FUND.

To December 28, 1915.

RECEIPTS.

Cash on hand Jan. 1, 1915	$678.88	
Interest on bank deposit.....	8.98	
		$687.86

EXPENDITURES.

Feb. 5. Paid John Koren	$67.00	
Oct. 9. Paid Rumford Printing Co. ..	200.00	
Nov. 19. Paid John Koren.....	251.50	
Balance on hand December 28, 1915 ..	169.36	
		$687.86

S. B. PEARMAIN,
Treasurer.

AUDITORS' REPORT.

We have audited the Anniversary Fund accounts of the treasurer of the American Statistical Association for the year ending December 28, 1915, and find the statements correct and the expenditures properly vouched.

Respectfully submitted,
LEROY D. PEAVEY,
MARTIN J. SHUGRUE,
Auditing Committee.

Digitized by Google

On motion, duly made and seconded, it was voted to accept the reports of the treasurer and the auditors, and to publish the same in the proceedings of the meeting.

It was moved and carried that a statement of the Anniversary Fund, held by the treasurer, be printed in the proceedings of the meeting.

The report of the Committee on Library was presented by Edmund E. Day. It was moved and carried that the Committee on Library be instructed to make a thorough investigation of the status of the relation between the library of this Association and the Boston Public Library and to report its findings and recommendations at the next annual meeting of the Association.

Julius H. Parmelee presented the report of the Committee on Constitution and By-Laws. The report was accepted, and after some discussion of the separate articles the following Constitution and By-Laws were adopted by unanimous vote:

Constitution.

Article I. This Association shall be denominated the American Statistical Association.

Art. II. The objects of the Association shall be to collect, preserve, and diffuse statistical information in the different departments of human knowledge.

Art. III. The membership of the Association shall consist of regular members, fellows, and honorary members.

Art. IV. Regular members shall be persons interested in statistics, elected by vote of the Board of Directors, on nomination by one or more members.

Art. V. Fellows shall be statisticians of established reputation, elected by unanimous vote of the Committee on Fellows, provided for in Article VI, who shall canvass the field at least once a year and announce elections, if any, at the annual meeting of the Association. Fellows shall have all the privileges of regular members, including the right to hold office and to vote. The number of fellows shall not at any time exceed one hundred.

Art. VI. The Committee on Fellows shall consist of five fellows, and shall be appointed by the President. The first

Digitized by Google

appointment under this article shall be made within three months after the adoption of this Constitution, and in the order determined by the President shall be for one, two, three, four, and five years, respectively. Subsequent appointments shall be made by the President at the annual meeting of the Association and for terms of five years each, but no member of the Committee may serve two terms in immediate succession.

ART. VII. Persons distinguished in the field of statistics or of public service may be elected to honorary membership by unanimous vote of the Board of Directors. Honorary members shall be entitled to all the privileges of regular members of the Association, except that they may not hold office or vote.

ART. VIII. The officers of the Association shall be a President, three Vice-Presidents, a Secretary, a Treasurer, a Librarian, an Editor, and three Counsellors. These officers shall be elected every year, beginning with the first annual meeting succeeding the adoption of this Constitution, in the manner prescribed by Article IX and shall take office at the close of the annual meeting at which elected. The President, Vice-Presidents, Secretary, Treasurer, and Counsellors shall form a Board of Directors for the government of the Association, three of whom shall constitute a quorum at any meeting regularly convened.

ART. IX. The President, Vice-Presidents, Secretary, Treasurer, Librarian, Editor, and Counsellors shall be elected in the following manner: A nominating committee of three members shall be appointed by the President, the personnel of which shall be announced in the Quarterly Publications of the Association at least three months before the annual meeting. It shall be the duty of the nominating committee to make nominations for officers to be elected at the annual meeting, and to notify the regular members and fellows by mail of its nominations at least thirty days before the annual meeting. Twelve or more regular members or fellows may make nominations for any office by submitting signed petitions to the nominating committee at any time after the appointment of the committee and up to the time of the annual meeting, but this shall not prevent the committee from making additional nominations for the same office.

Digitized by Google

If no nominations by petitions are submitted for any office, the nominee of the nominating committee may be elected at the annual meeting by *viva voce* vote. If there are additional nominations by petition for one or more offices, it shall be the duty of the nominating committee to announce such nominations, the election for such office or offices shall be by ballot, and the President shall appoint three judges of election who shall count the ballots, shall decide any question that may arise as to the validity of any ballot, and shall declare the results of the election. The candidate for President, or for Secretary, or for Treasurer, or for Librarian, or for Editor, receiving a plurality shall be declared elected, while the three candidates for Vice-President and the three candidates for Counsellors receiving the highest number of votes shall be declared elected.

ART. X. Annual and occasional meetings shall be held at such times and places as the Board of Directors may designate. Any vacancy in office resulting from death, resignation, or other cause shall be filled by election at the next meeting of the Association.

ART. XI. Amendments to this Constitution may be proposed by any member of the Association at the annual meeting, and if supported by at least one third of the members present and voting shall be submitted to the next annual meeting, where a two thirds vote of all members present and voting shall be required for adoption.

BY-LAWS.

1. At least once every two years the Secretary, acting in behalf of the Association, shall publish the Constitution and By-Laws and a list of members, and shall supply a copy to each member.

2. Each regular member shall pay annual dues of two dollars, or twenty dollars at some one time, and each fellow shall pay annual dues of five dollars, or fifty dollars at some one time.

3. The Board of Directors, as soon as practicable after their election, shall appoint two standing committees of three members each, namely, on Finance and on the Library; Five standing committees of five members each, namely on Federal

Digitized by Google

Statistics, State Statistics, Municipal Statistics, Business Statistics, and International Statistics. These committees shall serve for one year, or until their successors are appointed.

4. The Board of Directors, as soon as practicable after their election, shall appoint four associate editors, to serve with the Editor as a committee on publications, and to edit the publications of the Association. The associate editors shall serve for one year, or until their successors are appointed.

5. The Board of Directors, as soon as practicable after their election, shall appoint two assistant secretaries, who shall serve for one year, or until their successors are appointed.

6. Amendments to these By-Laws may be proposed at any annual meeting of the Association and if approved by a two thirds vote shall be declared adopted.

The Nominating Committee, consisting of Irving Fisher, Wesley C. Mitchell, and Frank R. Rutter, reported the following nominations for office in the Association during the ensuing year: President, Charles P. Neill; Vice-Presidents, Charles F. Gettemy, Joseph A. Hill, Charles H. Verrill; Librarian, Horace G. Wadlin; Treasurer, S. B. Pearmain; Secretary, Carroll W. Doten; Counsellors, John Koren, E. Dana Durand, Allyn A. Young; Editor, William B. Bailey.

On motion the report of the Nominating Committee was accepted, and it was voted to instruct the secretary to cast one ballot for the list as presented. This having been done, they were duly declared elected.

Meeting adjourned.

CARROLL W. DOTEN,
Secretary.

MEETING OF THE BOARD OF DIRECTORS.

A meeting of the Board of Directors was held Thursday evening, December 30, and the following appointments were made in accordance with the provisions of the new Constitution and By-Laws of the Association.

Assistant Secretaries—Robert E. Chaddock, New York, N. Y., and Henry J. Harris, Washington, D. C.

Digitized by Google

Associate Editors—Louis I. Dublin, Julius H. Parmelee, Carl J. West, and John Cummings.

Committee on Finance—Harvey S. Chase, chairman, Walter E. Weyl, and S. B. Pearmain.

Committee on Library—Roger W. Babson, chairman, Edmund E. Day, and Henry J. Harris.

Committee on Federal Statistics—John Koren, chairman, Thomas S. Adams, Leonard W. Hatch, Frederick L. Hoffman, and N. I. Stone.

Committee on State Statistics—E. Dana Durand, chairman, Robert E. Chaddock, Charles F. Gettemy, Henry J. Harris, and E. W. Kemmerer.

Committee on Municipal Statistics—LeGrand Powers, chairman, Frederick A. Cleveland, Fred C. Croxton, Louis I. Dublin, and Edward M. Hartwell.

Committee on Business Statistics—W. S. Gifford, chairman, Melvin T. Copeland, Julius H. Parmelee, H. S. Person, and I. M. Rubinow.

Committee on International Statistics—Royal Meeker, chairman, Roger W. Babson, Miles M. Dawson, Irving Fisher, and Wesley C. Mitchell.

Auditing Committee—Martin J. Shugrue and Ernest H. Gaunt.

CARROLL W. DOTEN,
Secretary.

Digitized by Google

REVIEWS AND NOTES.

Sweden—Historical and Statistical Handbook by Order of the Swedish Government. Edited by J. Guinchard. Second Edition, English Issue. Stockholm Government Printing Office, 1914. First part "Land and People," 781 pp. Second part "Industries," 758 pp.

In 1900 the first edition of this handbook under the editorship of Dr. Gustav Sundbärg was published in French on account of the Paris Universal Exposition. An edition in Swedish followed this in 1901 and in English in 1904.

In 1912 the present editor was commissioned to publish a new edition. In view of the Baltic Exhibition it was decided to publish a German issue and this was completed in May, 1914. It was originally planned to follow this with a Swedish issue but on account of the Panama-Pacific International Exhibition it was decided to publish an English issue first. This will be followed by the Swedish issue. The work under consideration is the second English edition.

The first volume dealing with Land and People is divided into five main divisions: Physical Geography, the Swedish People, Constitution and Administration, Education and Mental Culture, and Social Movements. To the average reader of these publications, the second and fifth of these chapters will probably possess the most interest. A large mass of statistical material is given upon moral conditions, criminality, customs and mode of life, and dwellings. This is well arranged and illustrated with maps and charts. Under social movements considerable space is given to labor questions and social politics. Such topics as Immigration, Unemployment, and Cost of Living are considered. The closing section of the book is devoted to a consideration of social movements. Under this heading we find the care of the needy and destitute, efforts in social hygiene, social rescue work, protection of children, eugenics, and the work of the Swedish National Anti-Tuberculosis Association. Two short sections are devoted to political economy and sociology and to statistics. The latter was written by K. A. Edin. It deals principally with the various statistical bureaus of Sweden and gives in some cases very brief biographies of the men who have been most prominent in statistical work in the country.

The second part discusses the natural resources of Sweden, agriculture, forestry, mining, manufacture, commerce, banking, and a general synopsis of trade and industrial legislation. All of the principal industries of Sweden are treated in a separate section. Charts are used in many cases to illustrate the growth of these industries. By means of well drawn maps, the deposits of ores and coal are shown. Other maps give the location of the railways, canals, and waterways of the kingdom.

Hundreds of illustrations are scattered throughout the volumes. No one can read this work without receiving a comprehensive impression of Sweden, her people, her resources, and her activities. The scheme of the

Digitized by Google

work was elaborate and the chapters are well balanced. The aim of the Riksdag has apparently been accomplished. The work of editing such a publication was by no means an easy one but it has been well done.

WM. B. BAILEY.

A SIMPLE METHOD FOR FINDING THE MOVING AVERAGE.

In the *Quarterly Publications* of the American Statistical Association for December, 1915, Mr. King, of the University of Wisconsin, offers a new method for obtaining the moving average of an historical variable. His method is certainly simpler than the long direct method, and also preferable to some short-cut methods. A method, however, which is still simpler than the one put forth by Mr. King has been used by the writer in statistical work in the University of Minnesota. It involves only three operations, namely, summation, subtraction, and division, whereas the other involves four.

1. Summate the items of the variable on the adding machine, taking sub-total for each item entered, including the first. The result is seen in the adduced representation of an adding machine slip which has been elaborated into a table. The sub-totals are marked with an asterisk. They may be numbered if convenient. The figures given are the average prices of cotton for the decade 1890–1899. The wave length is here for convenience taken to include only five items of the series.

Year.	Average Annual Price of Cotton in Cents per Pound. Sub-totals Marked with an Asterisk.	Five Year Group Totals.	Moving Average or Trend.
1890	8.6		
	8.6*		
1891	7.3		
	15.9*		
1892	8.4		
	24.3*		
1893	7.5		
	31.8*		
1894	5.9		
	37.7*	37.7	7.5
1895	8.2		
	45.9*	37.3	7.5
1896	7.3		
	53.2*	37.3	7.5
1897	5.6		
	58.8*	34.5	6.9
1898	4.9		
	63.7*	31.9	6.4
1899	7.6		
	71.3*	33.6	6.7

2. To obtain the total for each successive group of five items subtraction is necessary. Obviously, the total for the group of the first five items is the fifth sub-total. The total for the second group is obtained by sub-

Digitized by Google

tracting the first sub-total from the sixth; the total for the third group by subtracting the second sub-total from the seventh; and so on as far as we desire to go. In order to facilitate the process of subtraction, the most convenient device used by the writer has been to take a strip of cardboard or stiff paper cut to a size to cover the figures *intervening* between the subtrahend and the minuend. If the spacing is regular, it is necessary simply to slide the strip past one sub-total for each subtraction performed. The rectangle enclosing all the items between the first and sixth sub-totals indicates the use of the card for the first subtraction. With a little practice this subtraction can be done very rapidly. If the superior position of the subtrahend causes trouble, this may be avoided by leaving an extra space or two below each sub-total where the subtrahend may be written in, what is to the ordinary person, its conventional position.

3. To obtain the average, divide each group total by the number of items in the wave-length, in this case by five. The results are all shown on the slip, both the group-totals and group-averages being shown, the latter furthest to the right.

It is believed there is an advantage in thus placing the results of all three operations on the slip. For most purposes of publication or demonstration, the figures cannot be entered directly on the tables by the adding machine. From the slip they may be easily transposed. The slip may easily be made into table form if desired.

It is, of course, to be remembered that the average logically belongs not opposite the last item in the wave-length, but opposite the middle. But if students understand the meaning of the moving average, they will not meet with difficulty on this account.

It is very easy to verify the results. Simply summate the items of the last wave-length. If this sum is the same as the last group-total, the results are correct, except for the extreme improbability of compensating errors. If it is deemed necessary to verify the averages obtained by division, this may be done by summating the averages and the group-totals and dividing the sum of the group-totals by the number of items in the wave-length. The quotient should very closely approximate the sum of the averages.

JENS P. JENSEN.

University of Minnesota.

NOTE ON THE MOVING AVERAGE.*

I have been interested in reading the paper entitled "New Method for Computing the Moving Average," by Mr. Willford I. King, published in the December number of the QUARTERLY PUBLICATIONS of the American Statistical Association.

Mr. King proposes, by means of the adding machine, to add the yearly items, taking a sub-total after each item, to practically repeat the process,

* A similar note has been received from W. M. Persons, suggesting this improvement in the Method of Computing the Average.—ED.

Digitized by Google

suitably paste the two adding machine slips together, and by cross-subtraction obtain the remainders, which are shown in Column No. 3. These remainders being divided by seven, give the moving average.

I am writing to suggest that the process may be shortened and the cross-subtraction avoided by means of a simple device. For example, Mr. King shows the sum of the first seven items in Column No. 1, being the items for the years 1890–1896 inclusive. He then adds the item for the year 1897 and takes another total. If before taking this total he had also added in the complement of the item for the year 1890, he would have directly arrived at the second item in Column No. 3. If the adding machine has a capacity of nine figures, 100,000,000 will be one beyond the capacity of the machine. If the difference between 100,000,000 and any item is added, the result is the same as subtracting the item in question, for the carrying mechanism does not carry beyond the capacity of the machine.

The inclosed slip shows the calculations according to the plan I suggest for the first few years. It occurred to me that the use of the adding machine which I suggest may not be generally thought of, although I had supposed, before reading Mr. King's article, that the use of the complement for the purpose of accomplishing subtraction was in fairly general use.

Year.	Central Year.		Year.	Central Year.	
1890		9,203		1895	62,436
1891		8,280	1899		13,621
1892		9,157	1892		9,990,843
1893		7,125		1896	66,900
1894		6,658	1900		13,789
1895		9,446	1893		9,992,875
1896		8,623		1897	73,564
	1893	58,492	1901		15,878
1897		9,653	1894		9,993,342
1890		9,990,797		1898	82,784
	1894	58,942			
1898		11,774			
1891		9,991,720			

PERCY C. H. PAPPS.

Newark, N. J.

SUICIDE RECORD FOR 1914, BY FREDERICK L. HOFFMAN, THE SPECTATOR, NOVEMBER 25, 1915.

This article is Mr. Hoffman's twentieth annual survey of suicide in principal cities of the United States. In each of these studies special reference was made to the problem as it concerns the institution of life insurance. The present article ably reviews the literature of insurance practice on the suicide question, and recommends a thorough inquiry into

Digitized by Google

the subject with a sincere regard for the essentials of approved actuarial analysis of the available data.

We may well agree that in the experience of life insurance companies and according to the most trustworthy vital statistics, the suicide *rate* is on the increase; but it is still a matter of reserved reflection whether suicide itself has actually increased, with respect to the sharply changing characteristics of population and of registration practice. Suicide registration has been a particularly unsatisfactory element of our American vital statistics, and the observed increase in the rate over the past ten years may be deemed, in one sense, an index of improved registration rather than a measure of acceleration in the relative frequency of self-murder.

Mr. Hoffman has performed a distinct public service in publishing his annual surveys of suicide and homicide. His endeavors have added to the sum total of ascertainable facts on these two, of many, complex social phenomena.

E. W. KOPF.

IMPROVEMENT IN TRADE STATISTICS.

On February 9, Secretary Redfield of the Department of Commerce, addressed the annual meeting of the Chamber of Commerce of the United States in Washington, and explained the newly inaugurated system for compiling trustworthy export and import trade statistics.

In January, officials of the Chamber complained of inaccuracies in the published figures on exports for the year 1914. Mr. Redfield showed that the principal faults in trade statistics resulted from inadequate returns furnished by exporters and importers. The statistics are collected under laws of 1821 and the defects in the tabulations could well be ascribed to the limitations of this legislation.

New export regulations by the Secretaries of the Treasury and of Commerce correct the most flagrant abuse of failure to obtain complete statements for all shipments. No vessel may now clear from any port without a declaration being filed for every shipment. Furthermore, three men with statistical experience edit every entry before permits are issued. Import statistics may now be expected under more detailed and informing classifications. The introduction of modern tabulating machinery in the New York Custom House will shorten the time required for tabulation and will materially advance the accuracy of the published results.

E. W. KOPF.

Digitized by Google

STATISTICAL BOOKS ADDED TO THE BOSTON PUBLIC LIBRARY* DURING QUARTER ENDING DECEMBER 31, 1915.

PREPARED BY LIBRARY COMMITTEE, ROGER W. BABSON, EDMUND E. DAY, HORACE L. WHEELER.

Anuario de la América latina. Índice de industrias en español, alemán, francés, inglés, italiano y portugués. Directorgerente: Eduardo Riera Solanich. Año 2. Barcelona. 1914. Maps. 9318.a2

Babson, Roger Ward. Bonds and stocks. The elements of successful investing. Wellesley Hills. [1914.] Charts. Tables. 9332.6a34

—The future method of investing money. Boston. 1914. Chart. 9332.6a44

—The future of the railroads. Boston. 1914. Tables. Chart. 9385.973a120

Capper, Charles. The port and trade of London. London. 1862. Tables. 9387.9421

Chicago Traction Valuation Commission. Report on the values of the properties of the Chicago Consolidated Traction Company inside the city limits. [Chicago. 1910.] Maps. 9388.773a4

Gras, Norman Scott Brien. The evolution of the English corn market, from the twelfth to the eighteenth century. Cambridge. 1915. Map. Charts. [Harvard economic studies.] "Awarded the David A. Wells Prize for the year 1912–13." Bibliography, pp. 465–479. 9382.1142a5

Hooper, William Everett. Railroad accounting. New York. 1915. [Appleton's Railway series.] 9385.a33

International Institute of Agriculture. Bureau of General Statistics. Annuaire international de statistique agricole, 1911 et 1912. Rome. 1914. 9338.1a22

Leroy-Beaulieu, Pierre. Les impôts et les revenus en France, en Angleterre et en Allemagne. Paris. 1914. A study of internal revenue. 9336.27a25

Mallock, William Hurrell. Social reform as related to realities and delusions. An examination of the increase and distribution of wealth from 1801–1910. New York. 1915. 9330.1a68

McFarlane, John. Economic geography. London. [1914.] Maps. List of works consulted, pp. 527–529. 9338.a3

Pennsylvania. Department of Labor and Industry. Annual report of the Commissioner. 1st, part 1, 2. Harrisburg. 1915. Charts. Tables. 9331.0748a3

—Public Service Commission. Annual report, 1st. Harrisburg. 1915. Tables. 9381.09748

*The Library of the American Statistical Association is housed at the Boston Public Library and constitutes the major part of its statistical division.

Digitized by Google

Smith, Joseph Russell. Industrial and commercial geography. New York. 1913. Illus. Maps. Plans. Charts. Bibliography, pp. ix–xi. References are also appended to the different chapters. 9380.07a19

United States. Bureau of Labor Statistics. Foreign food prices as affected by the war. Washington. 1915. Tables. 9331.073a24.170

— —Hours, earnings, and conditions of labor of women in Indiana mercantile establishments and garment factories. Washington. 1914. Tables. 9331.073a24.160

— —Wholesale prices, 1890 to 1913. Washington. 1914. 9331.073a24.149

—Department of Commerce. Bureau of Foreign and Domestic Commerce. The pottery industry. Report on the cost of production in the earthenware and china industries of the United States, England, Germany and Austria. Washington. 1915. Tables. Charts. 9382.73a36.21

Utah. State Bureau of Immigration, Labor and Statistics. Report. 2d. 1913/14. Salt Lake City. 1915. Plates. Maps. Tables. 9317.92a2

Virginia. Special Joint Committee on Taxation. Report. 1914. Richmond. [1914.] Map. Tables. 9336.755a4

Indexing and Filing: A Manual of Standard Practice. By E. R. Hudders. The Ronald Press, New York, 1916. Pp. vii, 292. Price, $3 (postpaid).

This is a book that will appeal to statisticians as well as to business men. It is a revelation even to those who have some acquaintance with modern office devices and methods. At first thought it doesn't seem possible to write a book or even a thin pamphlet about so simple a matter as indexing and filing; but, when one remembers hours of time wasted in the vain attempt to find some paper or letter filed by the office boy or the new stenographer, he opens such a work as this with a chastened hopefulness and a predisposition to be grateful even for a little useful information.

It becomes apparent on reading this work that there is a science as well as an art of indexing and filing, and that the subject has a history and will soon have a respectable literature. The writer of this review long ago became convinced that every commercial school should give a course of instruction in this subject, and this book has confirmed that belief; it would serve as an excellent text-book for such a course.

The author describes and explains the many devices used in filing, from the simplest to the most elaborate. A large number of cuts supplement the text and make the explanations and instructions so clear that even the office boy could understand them. The theory and practice of indexing and cataloguing also receive adequate and illuminating treatment. Special chapters are devoted to purchase, sales, and credit records, filing in lawyers' offices, architectural filing, etc.

Digitized by Google

This book will fill a long-felt want. It will enable one who does not require anything elaborate or complex to plan and install his own filing system, and it will save much time and trouble in the training of clerical employees.

C. W. DOTEN.

Massachusetts Institute of Technology.

NOTE.

The following is an abstract of the recently issued population returns of British North Borneo for the year 1911, showing the racial distribution according to sex, and with the distinction of adults and children. The statement indicates an extremely complex population, and brings out the very suggestive fact that out of a total population of 208,183, only 355 are Europeans and 57 Eurasians:

DETAILS OF POPULATION THROUGHOUT BRITISH NORTH BORNEO.
CENSUS 1911.

	Adults.		Children.		Total.
	M.	F.	M.	F.	
Europeans	276	38	22	19	355
Eurasian	19	15	6	17	57
Chinese	19,354	2,845	2,157	1,646	26,002
Japanese	89	145	5	7	246
Siamese	11	1	1	4	17
Arabs	16	6	6	7	35
Soudanese	48	16	2	.....	66
Somali	.....	1	.....	.....	1
Philippinoes	166	51	46	28	291
Malays	980	298	203	131	1,612
Natives of India and Ceylon	741	60	64	37	902
Natives of Neth: East Indies	3,473	1,289	424	325	5,511
Natives of Sulu Archipelago	2,012	1,800	901	790	5,503
NATIVES OF BORNEO:					
Bajaws	6,756	6,935	4,521	4,375	22,587
Bruneis	2,148	1,965	1,456	1,308	6,877
Dusuns	26,814	28,141	16,939	16,057	87,951
Dyaks	356	22	58	69	505
Idahan	282	348	148	169	947
Illanun	429	574	303	335	1,641
Kedayan	685	731	509	530	2,455
Muruts	8,798	8,907	3,988	3,621	25,314
Orang Padas (Besayahs)	1,373	1,432	1,064	1,009	4,878
Orang Sungei	2,605	2,617	1,986	1,960	9,168
Tagals	706	842	207	157	1,912
Tambunwa	423	392	314	240	1,369
Tidongs	536	540	398	303	1,777
Tutongs	45	70	46	43	204
	79,141	60,081	35,774	33,187	208,183

F. L. H.

Digitized by Google

Digitized by Google

AMERICAN
STATISTICAL ASSOCIATION

NEW SERIES, No. 114. JUNE, 1916.

SOME PROBLEMS OF POPULATION GROWTH.*

By E. Dana Durand.

The greatest phenomenon of recent times is the present European war. That war is, in considerable measure, the outgrowth of present or anticipated pressure of population upon means of subsistence. The problems which are most seriously engaging the attention of the present federal administration are the maintenance of our rights as a neutral nation, the increase of armament, the expansion of export trade, and the improvement of rural credit facilities. All four of these problems arise largely from population growth. Though economists and sociologists very generally recognize that increase in population brings serious problems in its train, the same is by no means true of the great majority of the American people. Even though little that is new can be said on the subject, its vital character makes repetition and emphasis from time to time desirable.

One of the most striking features of modern history has been the rapid increase in the numbers of the white race, or what may be called the European stock. The white population of the earth approximately trebled during the nineteenth century. In Europe itself the increase has been great, but there has been an enormous overflow into other continents. Most of all has the European stock found room for expansion in the United States. But elsewhere in the Americas, in Australia, in South Africa, and in North Asia, it has also settled in considerable numbers. No equal period of time ever

*Presidential address at the annual meeting of the American Statistical Association, Washington, D. C., December 28, 1915.

Digitized by Google

witnessed such an amazing rate of population growth on such a large scale.

This phenomenon of the expansion of the world's white population has been the more remarkable because there occurred at the same time a marked advance in the standards of living of that population. Instead of feeling a growing pressure of numbers on natural resources, the European stock enjoyed increasing abundance for the average man. Prosperity reigned. People grew to laugh at the gloomy forebodings of Malthus as to the effects of population growth. But the past offers no guarantee for the future in this respect. The coincidence of enormous increase in numbers with a rising standard of living was due in part to rapid progress in methods of production both in agriculture and in other industries, but in at least equal measure it was attributable to that geographical expansion to which reference has been made. Obviously such geographical expansion must ultimately find its limit. Already, in fact, the point has been reached where further geographical expansion meets with increasing difficulty.

At the beginning of the nineteenth century, a large part of the fertile temperate regions of the earth was still virtually unoccupied. Discovery and improvements in transportation brought enormous rich areas suddenly into the service of the European stock. Hundreds of millions of acres of excellent land in the United States, in Canada, in the Argentine, and elsewhere were brought into cultivation by European settlers and their descendants. This new land for the most part was quite as productive as the best of Europe itself. Geographical expansion of agricultural production for a long time involved no descent to inferior lands with consequent diminishing returns to labor.

Within the most recent years, however, the extension of agricultural areas has begun to invade less eligible fields. The temperate regions of the globe, within which only has the white race found it possible to live in large numbers, offer no further prospect of such magnificent additions to agricultural resources as took place during the eighteenth and nineteenth centuries.

It is true that there are in the temperate zone enormous areas

Digitized by Google

of land which are but thinly populated, which are as yet unused or but slightly used for agricultural production. The areas in which there is any considerable density of population are but a small fraction of the total land surface of the earth or even of the land surface of the temperate regions. But the great bulk of these undeveloped lands is seriously handicapped in one way or another. There are vast regions which are entirely desert and only a small fraction of which can ever be reclaimed by irrigation. Other great regions are too mountainous or rocky ever to contribute greatly to agricultural production. There remain, to be sure, hundreds, if not thousands, of millions of acres which are more capable of agricultural use, but for the most part even these are decidedly inferior to the better parts of Europe and America. They suffer from partial aridity, from cold, from lack of drainage, from poverty of soil, or from unfavorable configuration. Very little of the unoccupied area of the temperate zone can be compared for a moment with the splendid Mississippi valley, for example. Broadly speaking, in that zone labor of a given grade of efficiency will hereafter produce less per man with each further geographical extension of agriculture. Moreover, apart from the question of production, the comfort of human life in these areas remaining for settlement will in general be much lower than in present regions of large population.

Consider for a moment the characteristics of some of the largest temperate areas which remain undeveloped. First of all, the United States. Only one fourth of the total land area of this country is under cultivation, and less than half is in farms at all. Nevertheless, the unused or slightly used lands are either permanently deficient in some of the qualities essential to the best agricultural results, or they require great investment of capital to make them satisfactory agents of production. In fact, such geographical extension of agriculture as has taken place in this country in the last decade or two has for the most part been under distinctly unfavorable conditions.

From 1900 to 1910, 64,000,000 acres were added to the improved farm land of the United States—an increase of 15 per cent. Of this total, 54,000,000 acres, or fully five sixths, were

Digitized by Google

west of the 95th meridian; that is, west of the first tier of states on the further side of the Mississippi river. The great bulk of this new uncultivated land was in fact west of the 100th meridian. That meridian marks substantially the limit of normal humidity. West of it, irrigation or dry farming is in most cases necessary to make land productive. Irrigation, save in a few favored localities, requires very heavy investment of capital; the results of the more recent projects have not been altogether encouraging. Dry farming gives far less yield per acre per year and usually less production per man than farming under conditions of normal rainfall. The increase of our agricultural production through geographical extension during recent years has been accomplished only at increasing unit cost.

Much is expected of the enormous area of Canada. Great areas in the west and in the north of that country can unquestionably be brought under cultivation. But in much of this territory the rainfall is not sufficient to assure the regular production of good crops, and in most of it the cold is a serious handicap both to production and to comfortable living. The life of the farmer in a region where for months the days are exceedingly short and the cold intense—often falling to 40 or 50 degrees below zero—can hardly be considered a thing of joy. Even more unfavorable are the conditions in the vast unsettled areas of Siberia. Turning to the southern hemisphere we find a great proportion of Argentina, of South Africa and of Australia cursed with drouth. In fact, it is evident that, were there anywhere in the temperate regions large areas of good land still available for settlement, the people of the United States would scarcely be extending their agriculture into regions so ineligible as those where lie the greater part of the recent additions to our own cultivated area.

Another evidence of the fact that high grade agricultural lands are no longer readily being added to the world's use is found in the great advance in the prices of farm products and of farm lands, particularly in the United States. Grounds appear for the suspicion that in the last decade or two the standard of living of the working classes in this country, other than the farm owners themselves, has ceased to rise, if it has

Digitized by Google

not actually fallen, and that this change in tendency is due essentially to the filling up of our territory. Everywhere we hear bitter complaints of the high cost of living. By this is meant, of course, higher cost of living, decreased buying power of income. Prices have advanced rapidly, but advance in prices *per se* does not mean higher cost of living. It is unequal advance for different classes of commodities, or disparity between increase of wages and that of commodity prices, which gives real ground for dissatisfaction. The general advance in prices has been due primarily to the increased production of gold. The general tendency of wages and salaries to lag behind commodity prices when the supply of money is augmented, is the source of much of the higher cost of living. But no small part of it, at least in the United States, is due to the fact that agricultural products have risen faster in price than other products.

The accompanying table brings out the situation. The figures represent average index numbers of wholesale prices, duly weighted, for 1912 and 1913, the base in each case being the average prices for 1890 to 1899. The year 1914, in which prices of farm products were much higher still, is not used because of the possibly abnormal conditions due to the war. The data are compiled from the latest bulletin of the Bureau of Labor Statistics on wholesale prices, in which the methods of computation have been revised and made more scientific.

Under the heading, "farm products," in this table are included a number not so classed by the Bureau of Labor Statistics itself but which are placed by it in the group of food products. These include eggs, beans, milk, rice, onions, and potatoes,—wholly unmanufactured products of strictly farm origin,—and also butter, cheese, prunes, currants, and raisins, which, though in a sense manufactured articles, have been subjected to but slight change and that often on the farm itself. Most of the articles classed as farm products in this table are used directly or indirectly as human food, but a few, such as cotton, hides, and hay, are used for clothing or other purposes.

The remaining articles classed by the Bureau of Labor Statistics as food products have been divided into two groups:

Digitized by Google

the first includes manufactured articles derived from the products of domestic agriculture, such as flour, bread, meat; the second includes fish, mineral food substances, and imported food products,—sugar, tea, coffee, cocoa, spices, etc.

AVERAGE INDEX NUMBERS FOR 1912–13; AVERAGE PRICES OF 1890–99 EQUAL 100.

Group.	Index.
Farm products, raw or very slightly manufactured	158
Manufactured foods from domestic farm products	142
Other foods (mineral, fish, or imported)	101
Cloths and clothing	120
Fuel and lighting	139
Metals and metal products	113
Lumber and building materials	149
Drugs and chemicals	111
House furnishing goods	90
Miscellaneous	118
All commodities	137

This table shows a far greater price advance for domestic farm products than for most other classes of commodities. Their average price for 1912 and 1913 was 58 per cent. higher than the average for 1890–1899. For manufactured foods derived from domestic farm products, the increase was 42 per cent. As might be expected from the rapid exhaustion of our aboriginal forests, the increase for the group of lumber and building materials approaches that for farm products.

Confirmation of this great advance in the prices of farm products in the United States is afforded by the Census of 1910. According to the reports of the farmers themselves, the average increase in the prices of their crops, duly weighted, was no less than 66 per cent. between 1899 and 1909.

A comparison of the movement of wages with that of retail prices of food yields similar results. Such a comparison can be made only with difficulty; but Dr. I. M. Rubinow, in a recent careful analysis of the data of the Bureau of Labor Statistics, has reached the conclusion that in the short period from 1907 to 1912 the purchasing power of an hour's wages, expressed in cost of foods, decreased about 10 per cent., and

Digitized by Google

that for a considerable number of years before 1907 there had been no advance in such purchasing power.*

The extraordinary advance in the prices of farm products has naturally caused a still greater advance in the price of land itself. Anticipating still further increases in earning capacity, holders of land have marked up its price by leaps and bounds. From 1900 to 1910 the average value of farm land per acre in the United States as a whole advanced by no less than 108 per cent. In some of the best agricultural states like Illinois and Iowa, farm land values rose even more rapidly.

No doubt to use the extraordinary relative advance in prices of agricultural products and of land in the United States since the nineties as a measure of the increase in pressure of population on agricultural resources in the area of European settlement generally, would involve exaggeration. The middle nineties were years of abnormally low prices for farm products in this country. The imperfect data available for European countries show much less disparity in movement between the prices of agricultural and those of other commodities than the data for the United States. However, the prices of farm products in this country in recent years are not merely higher than during the nineties; they are the highest—reckoned in gold—ever known. Moreover, whatever may have been the movement of prices elsewhere and whatever the increase in the production of agricultural staples in the world as a whole, the filling up of the lands of the United States and the rapid advance in prices of agricultural products here can scarcely fail to indicate the near approach of a time when, for the world in general, pressure of population on agricultural resources can no longer find ready relief in geographical expansion. The United States represents a large fraction of the temperate zone and conditions here cannot but reflect themselves elsewhere.

Thus far, in discussing the future geographical expansion of agriculture, reference has been made only to the temperate zone. While a considerable part of the tropical zone is desert,

* Since this address was delivered, Prof. H. P. Fairchild has shown with a considerable degree of conclusiveness that the purchasing power of wages in this country was no higher in 1908 than in 1890.

Digitized by Google

there are large areas which are fertile and capable of great agricultural production. Tropical Asia and Oceania are densely peopled,—so densely, in fact, that they have often suffered from famine,—but the tropical regions of America and Africa are sparsely settled. The question arises, therefore, whether a large further expansion of the white population of the world may not find its food supply in these regions.

It is a well known fact that hitherto the white race has found the climate of the tropics an all but complete bar to permanent residence. The inhabitants of the tropics consist almost exclusively of colored races or of persons of mixed blood. Save only in Asia and some of the adjacent islands, these colored inhabitants of the tropics have, whether by reason of the depressing influence of the climate or otherwise, failed to develop much efficiency or to subdue in any appreciable degree the exuberant forces of nature.

It is conceivable that scientific progress and large investment of capital may enable the white race so to modify conditions in the tropics as to permit its members to dwell there in large numbers. The clearing of forests and the drainage of lands would much mitigate the injurious effect of the tropic heat. Could a large population of intelligent people maintain its efficiency in the now thinly settled tropics, it could unquestionably make them produce abundantly. At the same time, it must be recognized that the difficulties are great and that, at least for a long time to come, the prospects of such white settlement of the tropics are scarcely promising.

It has been suggested by some that the solution of the problem of utilizing the fertile tropics is to be found in developing there a large colored population under the direction and supervision of the whites. The success of the Dutch in the management of Java is often cited as an evidence of the possibilities of large production through colored labor in the tropics generally. In order that such a colored population should work assiduously and produce a surplus for exportation to the temperate zone, something more or less approaching slavery would apparently be necessary. Labor in Java is virtually compulsory, though of course it is free from the abuses of complete enslavement. The white race can scarcely look

Digitized by Google

forward with entire satisfaction to the prospect of having its increase in numbers depend upon the exploitation of colored peoples in the tropical zone.

We must conclude, then, that geographical expansion of agriculture in the future, whether in the temperate or the tropical regions of the earth, is likely to confront far greater difficulties than during the eighteenth and nineteenth centuries and that such difficulties will constantly become more serious.

It is urged, however, that scientific progress will permit a great increase in agricultural production and thereby furnish the means for a continuance of the growth of population. No doubt much is to be expected from scientific progress. It will help to render available lands which are now unsuited to efficient agriculture. It will do much to increase the yield per acre of lands already in use. Moreover, the improvement in productive processes in non-agricultural industries may, by setting free labor to turn to the soil, serve to increase the output of food.

But have we any right to assume that such progress will be sufficiently rapid to permit population to grow at the present rate and still advance its standard of living? Is it not usually the case that the earlier inventions and discoveries in science, art, and industry are relatively more important economically than the later? Can scientific seed selection and plant breeding, proper rotation of crops, efficient maintenance of soil fertility, or all the other improvements in agricultural methods which are now being studied and inculcated accomplish more for agriculture in a given length of time than did the introduction of horse drawn and automatic machinery for planting, cultivating, and harvesting crops during the nineteenth century?

We are sometimes misled into undue optimism as to the possibilities of agriculture by knowledge of the fact that in certain parts of the world enormously greater yields per acre are obtained than is the case elsewhere. We know, for instance, that the yields of the standard grains are twice as high in western Europe as in the United States or Canada or Argentina. We know that in China and Japan the cultivation of a single

Digitized by Google

acre often supports a family. We are too likely to overlook the fact that a large part of the excess in productivity is due to greater application of labor and only a much smaller part, if any, to superior methods. The output of each tiller of the soil is usually less in regions where the production per acre is high than where it is low. Intensive agriculture constantly bumps up against the law of decreasing returns.

It may be, of course, that the improvement in methods of production will be not slower but faster in the future than in the past. We have, however, no right to assume that such will be the case. A nation, like an individual family, is rash which multiplies its numbers in the blind confidence that "the Lord will provide." It is easily possible for population to outstrip agricultural production. Should the area of European settlement ever become as densely inhabited as China or India, it could hardly fail to mean a much lower standard of living than we now enjoy.

If a continuance of the present rate of population growth of the European stock as a whole portends the possibility of lower living standards, how much more must this be true of countries already so densely populated that they must rely on other lands for much of the necessaries of life? Several European countries have reached this stage. Further growth in their population with maintenance of present living standards is possible only by increasing dependence on imported foods and raw materials.

A large part of the enormous increase in international trade in recent decades has been due to the exchange of manufactured products of such countries for the raw products of less densely settled regions. The great economic problem of England, Germany, Belgium, and even France is usually expressed as the problem of finding markets for their surplus manufactures. This is but a euphemistic manner of saying that their problem is to find outside sources of food and other essentials. The search for those to whom to sell is really the search for those from whom to buy.

If present rates of population growth continue, other countries must one by one be added to the ranks of those dependent on outside sources for necessaries of life. This process of

Digitized by Google

building up population on the strength of resources somewhere else cannot go on forever. When enough nations have come to rely on the outside world for means of subsistence, there will be no outside world. There is danger that the end of the rope may be reached with a certain degree of suddenness.

From what lands can any great increase in supplies of food and other necessaries for the leading manufacturing countries be expected? Certainly not from the already overcrowded countries like India or China. For some time to come the more thinly settled regions, despite the handicaps under which most of them suffer, will doubtless have increasing surpluses of agricultural products to export. But these regions are rapidly filling up. Sooner or later they will require most if not all of their agricultural products to support their own population. Increasingly, moreover, they will come to manufacture their own more advanced commodities and there will be a consequent reduction in the demand for imports of manufactured articles. Indeed, the capitalists of the great industrial nations are themselves hastening the extinction of export markets by investing capital in factories and other industrial enterprises in the more thinly settled countries as well as in countries like China and India. Some day the great manufacturing countries are likely to be thrown back upon their own agricultural resources—resources which could not even support their present population without a serious decline in the standard of living.

The increasing dependence of certain nations on outside sources of food and other necessaries is a fertile seed of international discord. It is largely because of such dependence that England and Germany are today grappling in desperate conflict. A nation that is primarily devoted to manufactures and that can maintain its standard of living only by exporting them largely in exchange for food and materials naturally seeks to secure to itself exclusive sources of supply or, in the more common phrase, exclusive markets for its products. It looks with jealousy and fear upon the growth of the trade of another industrial nation with such sources of supply. Pressure of population in a crowded nation results in constant and strenuous efforts to control markets. It is all but inevitable

Digitized by Google

that this industrial and commercial struggle should ultimately result in war.

Nations which find themselves dependent on outside sources for necessaries of life are naturally disposed toward great armaments. The navy of England, the army of Germany, represent the effort to protect their sources of supply. They result from the desire to permit a steady increase in population without a fall in the standard of living. Indeed, there is a vicious circle in the policy of these nations. Germany, for example, wishes greater population that she may have more soldiers; she wishes more soldiers that she may expand her foreign trade and her territory; she wishes more foreign trade and more territory to support an increasing population. The sphere of woman, says the German Kaiser, is expressed in the four K's,—*Küche*, *Kleider*, *Kirche*, and *Kinder*,—with emphasis on the *Kinder*.

In the dependence of England and Germany on the outside world lies, too, the explanation of those violations of previously established principles of international law which have aroused such bitter complaint from our own country and other neutrals in the present war. In days when each belligerent was able to provide itself from within with food and necessaries of civilian life, it was natural that the list of contraband should be confined essentially to articles designed primarily for military use. No important effect on the results of the war could be secured by a wider extension of the list of contraband. Now conditions have so changed that to shut off sources of supply for an entire population may mean victory over the enemy. Under these circumstances belligerents can scarcely be expected to refrain from seeking that object or from adopting hitherto unaccustomed and unsanctioned means to accomplish it. The old rules of international law are hardly likely to be obeyed when the whole economic situation of belligerents is new.

The United States is already approaching the gateway of the path which England and Germany are treading. She is becoming primarily a manufacturing nation.

The great decline in our exports of agricultural products and the great increase in our exports of manufactures are

Digitized by Google

striking evidences of the change that is going on. Prior to 1900 we had been exporting over one third of our wheat crop. For the three years ending 1912, the proportion was less than one eighth. American fresh beef, which had furnished a large fraction of the consumption of the United Kingdom at the beginning of the present century, wholly disappeared from foreign markets only a decade later. The exportation of cured meats fell off greatly. In the last two or three years the United States has itself been importing considerable quantities of fresh meat from Argentina. It is true that during 1914 and 1915 exceptionally favorable weather conditions resulted in very large grain crops in this country and that our exports of agricultural products, stimulated by the unusual demand, have again increased greatly. There is, however, no reason to suppose that this condition will continue.

Meantime our rapidly increasing population has devoted itself more and more to elaborative industries. Our exports of manufactures have increased by leaps and bounds. Most Americans boast of these changes. The American politician has become accustomed to gauge the progress of the nation by the growth of manufactures and of exports of manufactures. Up to a certain point it was in fact an advantage to our country to develop in this direction. So long as growth in manufactures served to free us from dependence on older nations for the more advanced commodities, so long as it served to increase the variety and raise the quality of articles of consumption, it was most beneficial. So long as expansion in export trade served to bring us a greater range of imports and to place at our service things which our own land is incapable of producing, that expansion was advantageous.

But that which is up to a certain limit a gain may become an injury when once the limit is passed. We are rapidly approaching the time when we shall find it necessary to export manufactured commodities in order to supply ourselves with those very essentials of life which we have been accustomed to produce in abundance. If population in the United States continues to increase at the present rate and continues, as at present, to devote itself more and more to manufacturing industries, we shall soon become dependent on the outside

Digitized by Google

world in the sense in which England and Germany are dependent. We shall be threatened as they are threatened with a decline in the standard of living due to the possibility, if not the probability, of the drying up of outside sources of supply. We shall be drawn into trade conflicts which will threaten our peace. We shall feel forced to maintain great armaments.

Our statesmen are already turning their attention toward policies which indicate a sense of impending pressure of population on agricultural resources. They are beginning to talk of the necessity of finding wider markets for our manufactured products if labor is to be kept steadily employed. We are urged to take advantage of the disturbance of the foreign trade of the warring nations to strengthen our grip on the markets of South America, of Africa, of Asia. Various measures of governmental aid in the extension of markets and in the transportation of products to them have been taken or are now under consideration. It is being urged that combinations and trusts which in internal business would be unlawful should be permitted and even fostered for export business.

I am not quarreling with any of these policies. I am merely urging that we give serious consideration to the conditions and tendencies which lie back of them. If we cannot change the direction in which we are going, we should at least know what that direction is. We should not pride ourselves in pursuing a path which leads into danger.

The disastrous effects of overcrowding of population are not likely, under the institutions and laws existing in our country and in most other countries of European settlement, to be equally felt by all classes in the community. On the contrary, as long as natural resources are allowed to remain without restriction in private ownership, growth of population tends greatly to increase the disparity of wealth. It plays into the hands of owners of land and other natural resources. It is those that have no share in the ownership of resources whose standard of living is threatened. It is a familiar truism that increase of population creates economic rent, monopoly value, unearned increment.

In the early days of the single tax agitation in the United States stress was laid chiefly upon the injustice of leaving in

Digitized by Google

private hands the unearned increment of urban land, of mineral land, of water power, of railroads, and of similar forms of property. It was maintained that the greater part of the value of agricultural land, at least in the United States and countries of similar character, was not inherent in the land itself, but was a creation of the labor and capital of the farmers which should, according to these exponents of the doctrine, be free from taxation. Within the last decade or two this situation has changed. Economic rent for farm land in huge amount has emerged in the United States. A new agricultural problem has arisen as the direct result of growth of population. Reference has already been made to the extraordinary increase in land values in the United States in recent years. Without increasing in any way his labor or his investment of capital, without adding in the slightest to the quantity of products raised, the owner of farm land in the better agricultural sections might see his income rise immensely by reason of the advance in the prices of agricultural products.

Fortunately in this country the great bulk of the agricultural land is held by relatively small owners and five eighths of it by those who actually till the soil. The enormous unearned increment of land value has thus been widely distributed. Yet to the still more numerous landless class it is no great comfort that several millions have shared in that increment.

A certain amount of the increment of land value may properly enough in a new country be allowed to go to the pioneers and their immediate offspring. It is a fitting reward for initiative, compensation for hardship, stimulus to development. But in many parts of the United States the time has already arrived when justice to the landless classes demands that a large part of any future increment in the value of farm land should, by taxation or otherwise, be taken for general public use. Confiscation of values already existing would doubtless cause undue hardship and be unjust to recent purchasers, but further increments in the value of already valuable lands should no longer go to any single class.

A result of the unearned increment of farm land in the United States is that a rapidly increasing proportion of the

Digitized by Google

land is no longer cultivated by owners. Tenancy has assumed large proportions and is steadily advancing. So long as farm land commanded but little economic rent the owner, unless his tract was decidedly large, had to contribute his own management and physical labor to obtain a living from it. Today in many cases he can afford to live in idleness, to reap what others sow and tend.

The figures as to tenancy are familiar enough. In the United States as a whole 26 per cent. of the farms were operated by tenants in 1880, 28 per cent. in 1890, 35 per cent. in 1900, and 37 per cent. in 1910. Even more significant, however, are the data for the northern sections of the country. The problem of tenancy in the South, profoundly serious as it is, is in part a relic of slavery days. In the East North Central division, the proportion of tenant farms increased from 20 per cent. in 1880 to 27 per cent. in 1910; in the West North Central, from 20 per cent. to 31 per cent. These are our best agricultural sections. In Illinois and Iowa, the banner farming states, tenants in 1910 formed 41 and 38 per cent., respectively, of the farm operators. In one Illinois county out of five more than half of the farms are tenant farms.

While in a good many cases present farm tenants are sons or sons-in-law of retired farm owners and may themselves later become owners, nevertheless a distinct tenant class is rapidly developing. The tenancy problem is essentially the result of population growth.

It is in no small measure the increase of tenancy that has given rise to the present agitation for better facilities of rural credit. Except in newer and poorer agricultural sections, farm owners already enjoy reasonably satisfactory conditions of credit. They can borrow all they want at fairly moderate interest rates. It is very commonly stated that the real problem in the better agricultural sections is to devise means by which the tenant, the hired farm laborer, the immigrant, or the city dweller can obtain enough credit, at sufficiently low rates, to enable him to acquire and equip land.

Unfortunately in large parts of our country no such means can be devised, for the problem is only half a credit problem. The most ideal credit facilities would not alone suffice to

Digitized by Google

enable the landless to buy high grade land. The price of the land is too high and will remain too high so long as the unearned increment goes to the owner. Anticipating further advance in the price of farm products, the owner of agricultural land holds it at a figure such that its present output at present prices must necessarily be insufficient to give living wages to the farm operator and a normal rate of interest on the value of the land. It is very generally the case that the present earning capacity of the higher priced lands is only 2 or 3 per cent. of the value at which they are held. The farm owner capitalizes the present earning capacity at say 5 or 6 per cent. and then doubles the capitalization in anticipation of future advance in earning capacity. The poor man cannot afford to be a land speculator. He cannot buy and wait for an advance in earning capacity. He must pay interest on the cost of his land out of current earnings. Better credit facilities alone would do little to help him.

In fact a probable result of a material reduction in interest rates on land loans would be further to advance the price of farm land. Owners would very likely capitalize earning capacity on the basis of the lower rate. The landless man would therefore be substantially in the same position as before. It was not without reason that the National Conference on Rural Credit held at Chicago in December, 1915, included in its resolutions a recommendation that, in connection with any new system of rural credit, steps should be taken to prevent it from causing further advance in land values and inuring to the benefit of land speculators.

Unless we properly attack the problem of unearned increment a further increase in farm tenancy in the United States is bound to occur. Not only so but it is highly probable that the ownership of farm land will become increasingly concentrated and the economic rent more and more go to limited numbers of persons. It has ever been the case that monopoly values tend to be absorbed by the relatively few. When our multimillionaires come to realize the possibilities of unearned gain through farm land ownership, they will more and more seek control over such land, and their enormous resources will enable them to acquire wide reaching control.

Digitized by Google

Only passing reference may be made to the effects of increase of population in creating other forms of unearned increment. In the last decade or two there has been in the United States an immense addition to the value of urban land; of deposits of iron ore, anthracite coal, the better grades of bituminous coal, petroleum, copper, and other minerals; of water powers, forests, and similar limited natural resources. The great bulk of this unearned increment of value has been permitted to remain in private hands, and unfortunately in a very much smaller number of hands than in the case of farm land. Further additions to the value of these resources are bound to occur with further growth of population.

In the case of railroads and to some extent in the case of other public service enterprises, public regulation has checked materially the unearned increment due to population growth or, to put it more accurately, has taken that unearned increment in whole or in part for the public generally. Charges have been reduced or have been prevented from advancing in proportion to increase in operating expenses due to the generally higher level of prices and wages. Had there been no such rise in operating expenses, the railroads, for example, would, on the basis of an unchanged level of rates, have seen their profits enormously swelled during the past 15 or 20 years. The volume of traffic has grown very much more rapidly than the additions to investment. The more or less accidental fact of general advance in the level of prices and wages has enabled the people to absorb the unearned increment of the railroad business merely by holding rates substantially to the previous level.

It seems to me of fundamental importance that we in the United States should radically change our attitude toward growth of population. So long as we had vast resources scarcely scratched, rapid growth in population was eminently desirable. The scattered settlers in a new and rich country find their well-being distinctly promoted by the growth of numbers. But the pride in population growth and the desire for it engendered by such primitive conditions tend to persist after the conditions have utterly changed, and undoubtedly exercise a considerable influence in actually stimulating in-

Digitized by Google

crease in population. Further growth at the present rate threatens to lower our standard of living, to increase our dependence on other lands, and to widen the already perilously wide gulf between rich and poor.

The rate of growth of population tends to adjust itself to resources. Where population is already overcrowded and the standard of living low, famine and disease must increase. In a population with a high standard of living a voluntary restriction of the birth rate tends to take place whenever pressure on resources is felt. Indeed, such restriction of birth rate may be practiced with deliberate intent actually to raise the standard of living.

In this country this salutary tendency is, unfortunately, checked in some measure by that unreasoning pride in population growth to which reference has been made. Too often, moreover, both here and elsewhere, it is opposed by rulers and by leaders of thought and sentiment. The underlying motive of those who favor a high birth rate is usually the desire to increase military strength. Sometimes it finds support on supposed religious or ethical grounds. Sometimes the propaganda for a high birth rate is directed only to the middle and upper classes, who do, in fact, too often restrict their numbers unduly. The trouble is that those for whom the incitement to multiply is not intended are more apt to be influenced than those at whom it is aimed. Indeed, the danger of undue population growth lies largely in lack of prudence in respect to birth rate on the part of the poorer and less intelligent classes in the community. The effect of a high birth rate among such classes in increasing population is less subject to check by a high death rate in advanced countries than in semi-civilized countries. Our modern feeling of sympathy and our public and private donations for relieving distress protect the less prudent against the consequences of their multiplication in numbers.

In our own country the natural growth of population is greatly augmented by immigration. Had there been no immigration from 1900 to 1910 our population would have increased about 14 per cent. As it was, it increased 21 per cent. Immigration accounted for one third of the addition to our

Digitized by Google

numbers. Moreover, of the natural increase in the population a disproportionate share was attributable to the presence of the foreign born, of earlier migration. With their lower standard of living they have larger families than the older native stock.

If the admission of multitudes of foreigners to our shores served to relieve the pressure of population in the countries whence they came, we might at least congratulate ourselves upon a philanthropic action in opening wide our doors. As a matter of fact, it seems likely that the possibility of sending surplus population elsewhere tends to stimulate the birth rate in those countries. The policy of the United States and of various other nations in failing to restrict immigration, or even encouraging it, probably increases the rate of population growth in the world as a whole. However that may be, it is hard to be patient with those who believe that further large immigration will be beneficial to our own country. We no longer have a great over-plus of unused resources clamoring for development. It is principally into the manufacturing industries that present day immigrants are entering. Hardly an eighth of the foreign born added to our population during the last census decade went upon the land; the great bulk of them crowded into our cities. The flood of immigration is simply hastening the time when we shall become dependent on the outside world for the very means of subsistence, when our standard of living will fall, when we shall be forceed into militarism, when the masses of our people will be virtually the serfs of the few owners of natural resources.

Digitized by Google

THE BUDGETS OF SMITH COLLEGE GIRLS.

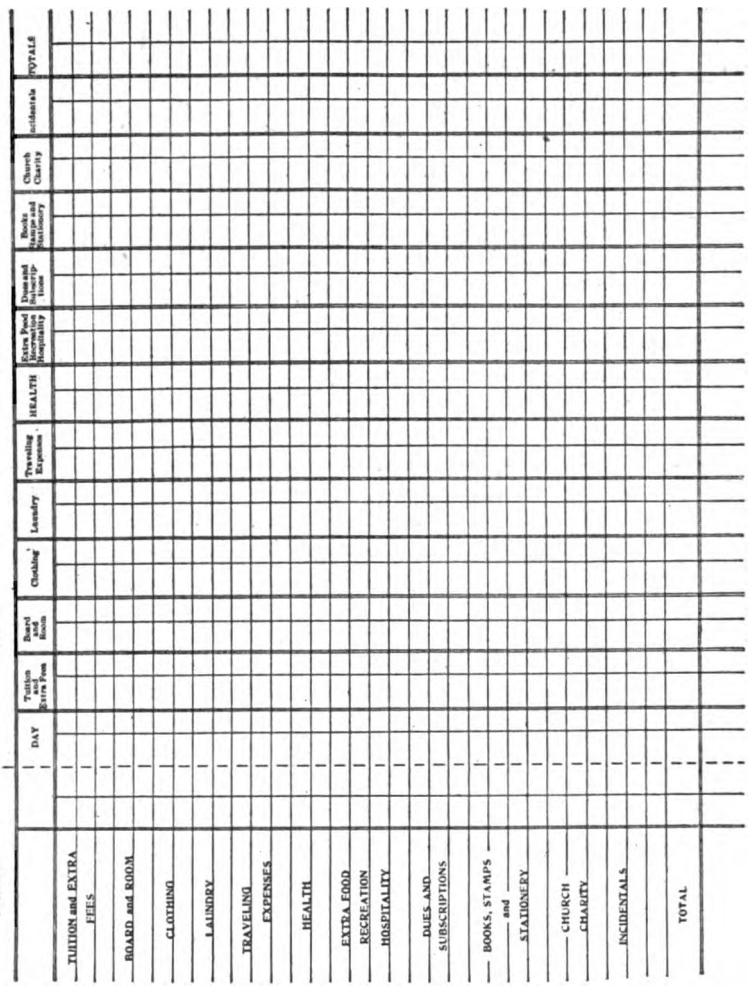
By F. Stuart Chapin,, Ph.D., *Associate Professor of Sociology, Smith College.*

During the academic year, September to June, 1914 to 1915, all students in Smith College were asked to keep expense accounts. The keeping of these records was entirely optional with the individual, and some 1,341 girls, out of a student body of 1,638, volunteered. About 900 accounts were discontinued in the course of the year, so that only 421 complete and accurate accounts were made the subject of analysis.

A committee of three, representing the administration, faculty, and student organizations, drew up the plan of procedure. It was decided to furnish each student with a standardized account book in the form of monthly sheets, providing for daily entries, attached to a stub. At the end of every month, the sheets with their daily entries complete, were turned in at the office of the Dean. The records were called in every month, since a shorter interval would have discouraged the students by making the work burdensome, and a longer interval endangered the accuracy of the recorded items. For similar reasons the expenditure categories were limited to the following eleven: tuition and extra fees; board and room; clothing; laundry; traveling expenses; health; extra food, recreation, and hospitality; dues and subscriptions; books, stamps, and stationery; church and charity; and incidentals.

The accounts were strictly confidential. Personal privacy was guaranteed by the following plan of administration: the student presidents of every college house distributed the books in September, and each month reminded the students of their share in the investigation; at the end of the month each student tore off the completed sheet and turned it in; since each account book and every one of its sheets bore the same number, the final assembling of the nine sheets of any one book was not a difficult matter. The stub, with its monthly totals, was retained by the student.

Digitized by Google



When the books were all in and assembled, every book was individually examined and carefully cross checked by an adding machine. The whole mass of accounts and all parts of the subsequent analyses were twice checked, each time by a different person. Innumerable minor errors in the student's addition were found to be compensating, but there were two constant sources of error which appeared in many of the books —failure to enter the second payments for board and room,

Digitized by Google

and for tuition. This necessitated the careful examination and perusal of every book and the addition of the omitted entry wherever it was fully justified. Since the object of the investigation was to determine the cost of the college year for the "average" student, additions had to be made to the accounts of scholarship students, to make up for the reduced entry for tuition caused by the stipend.

The representativeness of the 421 complete accounts was checked in four ways:

1. By determining the number of accounts from each academic class, and the percentage it bore to the total registration of that class. For example—there were 97 Freshmen accounts, or 16.7 per cent. of the class; 102 Sophomores, or 28.7 per cent. of the class; 138 Juniors, or 27.4 per cent. of the class; and 82 Seniors, or 26.7 per cent. of the class.

2. By determining the percentage of books from different types of dormitories. For example, it was found that representative large and small "Campus Houses" turned in complete, 50 per cent. and over of the books originally taken out.

3. A comparison of the total expenditures for the normal months of November and February from the 421 complete accounts, with similar totals obtained from all incomplete accounts. This comparison showed that the 421 complete accounts averaged $3.00 less than the incomplete accounts.

4. A comparison of the average expenditures of the 421 books, with the itemized expenditures of certain individual accounts, of students and alumnae, known to be accurate. (Comstock, A. L.—Undergraduate Budgets, *Smith Alumnae Quarterly*, Vol. VII, No. 2, pp. 84–86.)

These checks indicate that the completed accounts are fairly representative of budgets of the student body.

A total of $322,297.42 was expended by these 421 students during the academic year 1914–1915. The mean budget was, therefore, $765.55. Since this average tends to mask significant differences in expenditure which occur among different income groups, the individual totals are arranged by $50.00 class intervals in the following table:

Digitized by Google

TABLE I.

Total Expenditure.	No. Students.	Total Expenditure.	No. Students.
$350 to 400	4	$1,100 to 1,150	10
400 " 450	15	1,150 " 1,200	6
450 " 500	21	1,200 " 1,250	6
500 " 550	26	1,250 " 1,300	4
550 " 600	39	1,300 " 1,350	4
600 " 650	46	1,350 " 1,400	3
650 " 700	52	1,400 " 1,450	3
700 " 750	32	1,450 " 1,500	1
750 " 800	34	1,500 " 1,550	1
800 " 850	34	1,550 " 1,600	..
850 " 900	24	1,600 " 1,650	2
900 " 950	17	1,650 " 1,700	1
950 " 1,000	14	1,700 " 1,750	..
1,000 " 1,050	11	1,750 " 1,800	..
1,050 " 1,100	8	1,800 " 1,850	3
			421

Computing the modal average and the median from this table (King, W. I.,—Elements of Statistical Method, 1912, pp. 124, 129), the modal expenditure is found to be $670.51 the median $711.72, both considerably smaller accounts than the mean. The most representative average, however, is probably the mean, since our sample 421 accounts are undoubtedly drawn more largely from the economical and systematic students than from the extravagant classes. The skewing of the distribution over towards the lower income groups would appear to be an indication of this.

The mean expenditure for different items follows:

TABLE II.

Clothing	$76.27	Recreation	$35.09
Laundry	8.50	Dues and subscriptions	10.30
Travel	58.57	Books, stamps, stationery	26.08
Health	11.74	Church and charity	5.24
Incidentals	43.19		

But Table II has the disadvantage of masking significant differences found among different income classes. Table III throws considerable light upon the actual cost of different items and the proportion of total income expended for them.

Digitized by Google

TABLE III.

ITEMIZED EXPENDITURE BY INCOME CLASSES.(a)

Items.	$850–$900 Low.		$600–$800 Medium.		$800–$1,000 Liberal.		$1,000–$1,900 Very Liberal.	
	Average.	Per Cent.	Average.	Per Cent.	Average.	Per Cent.	Average.	Per Cent.
Board and room (b)	$251.00	48.6	$311.00	45.0	$362.00	41.2	$448.48	36.8
Clothing	11.09	2.1	46.17	6.6	90.62	10.3	216.67	17.7
Laundry (c)	1.61	.3	6.69	.9	10.82	1.2	19.23	1.5
Travel (d)	22.75	4.4	40.56	5.8	78.81	8.8	135.85	11.1
Health	5.47	1.0	10.35	1.5	14.95	1.7	19.38	1.5
Recreation	15.32	2.9	32.64	4.7	42.90	4.8	62.17	5.1
Dues and subscriptions	6.61	1.2	9.95	1.4	11.52	1.3	17.48	1.4
Books, stamps, stationery	16.73	3.2	25.27	3.6	30.28	3.6	37.98	3.1
Church and charity	2.69	.5	4.81	.6	6.01	.6	9.60	.7
Incidentals	19.36	3.7	36.22	5.2	58.28	6.6	81.15	6.6
Totals	$513.28		$689.66		$878.58		$1,218.60	
No. students per class	100		145		81		58	

(a) The table is based on 384 account books. Out of the 421 complete books, 37 had to be excluded in the computation of average expenditure for items because of blanket monthly totals. These totals were apparently accurate when checked with totals from itemised books, but as there was no way to distribute them among the items, the books had to be thrown out in computing the average itemised expenditures.

(b) Board and room is at the flat rate of $300.00 for the academic year in all campus houses, with the exception of Tenny and Lawrence House. In the latter houses, the charge is $200.00 and less. In the larger houses there are a few expensive suites which rent at from $400.00 to $500.00 for board and room. The charge for tuition is $150.00 for the academic year. But with extra fees added, tuition averages about $165.00. Grand totals at foot include tuition and extra fees.

(c) The item for laundry is small because a certain amount of laundry work is included within the charge for board and room in regular college dormitories, and also because of the custom which exists among off-campus students of sending their laundry home.

(d) Traveling expenses include to and from college during the Christmas recess and the Spring recess, but exclude expenses during the summer vacation.

If we select the itemized expenditures of the medium budget group as representing the expenditure scheme of the "average" girl, interesting comparisons may be made with the three other groups.

The girls with low budgets spend less than one fourth as much money and less than one third the "average" proportion of their income for clothing and laundry. The other expenditures of this class, especially those for travel, recreation and incidentals, are less than the average. Necessities, board and room, tuition and extra fees, absorb a relatively larger percentage of their more limited incomes than is the case with other classes.

The two liberal groups spend an increasingly greater proportion of their incomes for clothing and travel. Aside from

Digitized by Google

the variation in proportion of income expended for travel, clothing is the item most reduced by the economy of the low budget group and most increased by the extravagance of the liberal groups. The actual figures show that 99 low budget girls spent a total of $2,207.27 for clothing, as compared with $6,603.20 by 143 medium budget girls, $7,339.88 by 81 liberal budget girls, and $12,566.60 by 58 very liberal budget girls.

In connection with the expense schedule of the low budget group, it is interesting to examine the expenditures of 30 girls in Tenney and Lawrence Houses, college dormitories. At Tenney House, a girl may lodge for from $35.00 to $70.00 a year. At Lawrence House, where each girl takes care of her own room and gives approximately one hour a day to the work of the house, a total charge for board and room is $200.00. Lawrence House is a successful experiment in coöperative housekeeping.

TABLE IV.

ITEMIZED EXPENDITURE OF 30 GIRLS IN TENNEY AND LAWRENCE HOUSES.

Items.	Average Expenditure.	Per Cent. of Total.	Range.
Board and room	$180.75	37.0	$103.30–$263.00
Clothing	27.45	5.7	5.22– 101.80
Laundry	1.86	.3	.00– 4.61
Travel	23.35	4.9	1.06– 69.28
Health	8.89	1.8	.00– 66.91
Recreation	17.23	3.6	7.08– 71.52
Dues and subscriptions	7.78	1.6	1.89– 16.77
Books, stamps, stationery	19.32	4.0	5.47– 42.51
Church and charity	3.48	.7	.60– 14.90
Incidentals	20.04	4.2	1.24– 46.26
Totals	$476.51	63.8(a)	$377.27–$609.83

(a) excludes tuition and extra fees.

On the whole, these 30 girls appear to live rather better than the 99 girls of the low budget group of Table II. Since they save, on the average, about $70.00 a year on board and room, they are able to spend in spite of their smaller incomes a greater amount of money on every item in the list than are the girls of the low budget group. It is quite evident, therefore, that these low priced college dormitories abundantly justify their existence and are of great benefit to the girl of very moderate income.

Digitized by Google

A classification of mean expenditures of the entire group of students on the basis of necessities, pleasure and books, as adopted by Professor William B. Bailey ("Personal Budgets of Unmarried Persons," *Yale Review*, Vol. X, pp. 70–83), shows 87.4 per cent. for necessities, 8.2 per cent. for pleasure, and 4.2 per cent. for books.

TABLE V.

I. Necessities	87.4	II. Pleasure	8.2
Board and room	55.0	Dues	1.6
Clothing	12.4	Church and charity	.8
Laundry	1.3	III. Books, stamps, and stationery	4.2
Travel	9.5		
Health	1.9		
Incidentals	7.0		

(Payment for tuition and extra fees excluded)

The stability of certain expenditures in proportion to income is an interesting contrast to the variation in other items. Thus an examination of Table III shows that the proportionate expenditure for health, dues and subscriptions, books and stationery, and for church and charity varies but little whatever the income. The proportionate expenditure for recreation and incidentals remains fairly constant for the three upper groups. It is apparent, therefore, that these items constitute a rather stable element in the standard of living of the group of college girls studied.

The significance of this fact is further brought out by Table VI, in which the items are arranged in order of their proportionate importance by the income classes of Table III.

TABLE VI.

EXPENDITURE ITEMS BY ORDER OF IMPORTANCE.

	Low $350–$600.	Medium $600–$800.	Liberal $800–$1,000.	Very Liberal $1,000–$1,900.
1.	Board and Room	Board and Room	Board and Room	Board and room
2.	Travel	Clothing	Clothing	Clothing
3.	Incidentals	Travel	Travel	Travel
4.	Books	Incidentals	Incidentals	Incidentals
5.	Recreation	Recreation	Recreation	Recreation
6.	Clothing	Books	Books	Books
7.	Dues	Health	Health	Health
8.	Health	Dues	Dues	Laundry
9.	Church and charity	Laundry	Laundry	Dues
10.	Laundry	Church and charity	Church and charity	Church and charity

Digitized by Google

The median item appears to be recreation and it holds fifth place in all classes. The three upper classes are in agreement for the first seven items and church and charity. This table indicates that the entire group is quite homogeneous.

The present study would be incomplete without some mention of its relation to the principles of Engel's law. Table VII is therefore presented to show the bearing of the Smith College figures upon this law. It will be observed that the trend of the figures in the table is in substantial agreement with the American experience of itemized expenditure by different income classes: the greater the income the smaller the percentage outlay for subsistence; the greater the income the larger the percentage expended for clothing; and the greater the income the larger the percentage expended for sundries (recreation and incidentals).

TABLE VII.

SMITH COLLEGE FIGURES BY ENGEL'S GROUPING.

Items. Per Cent.	Low $350–$600.	Medium $600–$800.	Liberal $800–$1,000.	Very Liberal $1,000–$1,900.
Board and room.....	48.6	45.0	41.2	36.8
Clothing.....	2.1	6.6	10.3	17.7
Sundries.....	6.6	9.9	11.4	11.7

A study of expenditure by academic classes, Seniors, Juniors, Sophomores, and Freshmen, shows that in the order named the percentage spent for room and board diminishes. Juniors show the largest total expenditure for the year and the largest percentage spent for clothing, laundry, and recreation (the Junior "Prom" causes the swelling of these items). Seniors spend the largest percentage of any class for books.

Considering the facts summarized in the foregoing tables, it appears that Smith College students, as measured by the sample 421 account books, are a homogeneous average group of American girls.

Digitized by Google

A COMPARISON OF THE RELATIVE HEALTHFULNESS OF CERTAIN CITIES IN THE UNITED STATES BASED UPON THE STUDY OF THEIR VITAL STATISTICS.

By G. E. Harmon, M.D., C.P.H., *Department of Hygiene, Western Reserve University.*

In making a comparison of the relative healthfulness of several cities, the whole picture of the sanitary conditions must be obtained; in other words, the problem must not be approached from a single viewpoint. Take, for example, two communities, one having a high typhoid rate and a low infant mortality rate, the other having a low typhoid rate and a high infant mortality rate; if either the typhoid or the infant mortality rate is taken as the only criterion, are we able definitely and impartially to decide which of the two is the more healthful? Decidedly not, for we have approached the problem from one point of view and have omitted the measurement of other factors which go to make up the total health of a community. It is necessary, therefore, to make use of indices measuring as many as possible of the factors going to make up the total picture of the health of the community.

This, of course, is somewhat idealistic, and for practical purposes is not possible except perhaps by the use of life tables, which probably furnish the best single measurement of the health of a community. Many health factors are taken into account in the construction of a life table. This method of comparison has the drawback that life tables require a great deal of time and technical knowledge for their construction. In this paper, accordingly, the ideal method of comparing communities by the use of life tables will not be employed, but a few of what would appear to be the most valuable of the other indices will be considered.

The following cities will be considered: Baltimore, Boston, Buffalo, Cincinnati, Cleveland, Denver, Los Angeles, Minneapolis, New York, Omaha, Pittsburgh, San Francisco, and St.

Digitized by Google

Louis. A period of 15 years, 1900 to 1914, will be covered, for it seems to be advisable to include a sufficient number of years to avoid the fluctuations in the yearly rates due to other causes than actual health work. Because of these little understood variations, it is hardly fair to judge a city by one year's experience. It is the steady general tendency in health conditions resulting from organized purposeful health work which we wish to measure, and this can best be done by studying the vital statistics for a number of years.

For the first 13 years considered, the material has been obtained from the annual Mortality Statistics published by the Bureau of the Census. The data used for obtaining the figures for the last two years were furnished me by the health officers of the various cities, or were obtained from the annual health reports of these cities. In one or two instances certain figures for 1913 were obtained from the mortality statistics for that year instead of the annual reports of health officers.

In calculating rates, midyear estimates of the population were used. Whenever it was found necessary to estimate the population, the arithmetical method was employed. The estimates of population for 1910, 1911, 1912, and 1913 were taken from Bulletin 122 of the Bureau of the Census. Of course these estimates for the last 4 years of the period under consideration may be too low or too high, but this cannot be determined until the next census is taken. Thus the rates for the last 4 years may be slightly inaccurate; in some cases the rates will be too low and in others, where the population has grown more rapidly than has been estimated, the rates will be too high. The probability is, however, that these errors are small and of but little consequence. Then, too, the rates used are as likely to be correct as any that can now be obtained. The calculations for this paper were made by the use of a slide rule, as this method was deemed to give sufficient accuracy.

The Crude Death Rate is frequently used as an index in judging the sanitary condition of a community. It is not an altogether accurate index, for it is influenced by or depends upon a number of factors which do not come within the con-

Digitized by Google

trol of health authorities and which may vary a great deal in two places. The age and sex distribution of the population are two important factors which greatly influence the crude death rate. As these factors have nothing to do with the healthfulness of a given place, an index which is influenced by them is not altogether satisfactory. The crude death rates, however, give some general indications which may be of value, and are probably a fairly accurate measure of the changes over a period of years in a given community, for here the factors of age and sex distribution do not vary much from year to year.

The following table gives the average crude death rate per 1,000 population for the entire period, the average rate for the first three and the last three years of the period, and the decrease or increase between these last average rates:

TABLE I.

CRUDE DEATH RATES FROM ALL CAUSES.

City.	Average Rate, 1900–1914.	Average Rate, 1900–1902.	Average Rate, 1912–1914.	Increase or Decrease.
Baltimore	19.3	20.4	18.0	−2.4
Boston	17.9	19.7	16.3	−3.5
Buffalo	15.5	14.9	15.4	+0.5
Cincinnati	18.7	18.6	16.5	−2.1
Cleveland	14.5	16.0	13.6	−2.4
Denver	16.6	18.5	13.7	−4.8
Los Angeles	15.5	17.0	14.2	−2.8
Minneapolis	10.8	11.0	11.3	+0.3
New Orleans	21.9	23.6	20.2	−3.4
New York	17.2	19.6	14.2	−5.4
Omaha	12.8	12.0	13.4	+1.4
Pittsburgh	18.3	19.9	16.2	−3.7
San Francisco	17.7	20.7	15.6	−5.1
St. Louis	16.3	17.8	15.0	−2.8

It is evident from a study of this table that the rates in nearly all the cities have decreased in the last 15 years. The decrease is especially large in the case of New York, Denver, and San Francisco. It will be observed that the rates have increased for Buffalo, Minneapolis, and Omaha. In the case of Buffalo, a city which is growing very rapidly, this is probably due to the fact that the estimates of the population are too low. The same reason may explain the increase in the rates of the other two cities, or the increase may be due to imperfect registration of deaths during the early part of the period being

Digitized by Google

considered, which would tend to make the rates for the early years too low and thus prevent the cities from showing a decrease.

In rating cities as to their relative healthfulness it would hardly be fair to arrange them according to the percentage of decrease in their death rates, for a city with a low rate to start with would not have a possibility of reducing it equal to that of the city which started with a high rate. Moreover, as we are especially interested in the conditions obtaining at the present time, the average rates for the entire period and for the first three years are not satisfactory for rating purposes. We, therefore, will take the average rates for the years 1912–1914 as the basis in comparing the cities. We assume that the lower this rate, the more healthful is the city. The order of the cities in regard to their healthfulness, beginning with the one having the lowest rate, would then be as follows:

(1) Minneapolis	11.3	(8) Buffalo	15.4
(2) Omaha	13.4	(9) San Francisco	15.6
(3) Cleveland	13.6	(10) Pittsburgh	16.2
(4) Denver	13.7	(11) Boston	16.3
(5) New York	14.2	(12) Cincinnati	16.5
(6) Los Angeles	14.2	(13) Baltimore	18.0
(7) St. Louis	15.0	(14) New Orleans	20.2

Standardized Death Rates. As has already been indicated, the crude death rate is markedly influenced by the age and sex distribution of the population. Since in the standardized death rates the variable factors of age and sex distribution have been eliminated, they should furnish a more reliable index of sanitary conditions in comparing communities than the crude death rates. The standardized rates presented in this paper were computed by the indirect method as described in the Seventy-fourth Annual Report of the Registrar General of Births, Deaths, and Marriages in England and Wales, 1911.

Digitized by Google

TABLE II.

STANDARDIZED DEATH RATES.

City.	Average Rate, 1900–1914.	Average Rate, 1900–1902.	Average Rate, 1912–1914.	Increase or Decrease.
Baltimore	20.6	21.8	19.2	−2.6
Boston	19.1	21.0	17.4	−3.6
Buffalo	17.1	16.4	17.0	+0.6
Cincinnati	20.3	20.2	17.9	−2.3
Cleveland	16.2	17.9	15.2	−2.7
Denver	18.8	21.0	15.6	−5.4
Los Angeles	16.3	17.9	15.0	−2.9
Minneapolis	12.5	12.7	13.1	+0.4
New Orleans	24.3	26.1	22.4	−3.7
New York	19.7	22.5	16.3	−6.2
Omaha	14.9	13.9	15.5	+1.6
Pittsburgh	20.7	22.5	18.3	−4.2
San Francisco	20.1	23.5	17.7	−5.8
St. Louis	18.4	20.1	17.0	−3.1

On the basis of the standardized rates for the period of 1912–1914, the cities now have the following order:

(1) Minneapolis	13.1	(8) Buffalo	17.0
(2) Los Angeles	15.0	(9) Boston	17.4
(3) Cleveland	15.2	(10) San Francisco	17.7
(4) Omaha	15.5	(11) Cincinnati	17.9
(5) Denver	15.6	(12) Pittsburgh	18.3
(6) New York	16.3	(13) Baltimore	19.2
(7) St. Louis	17.0	(14) New Orleans	22.4

It will be noticed that the order of the cities is not the same as in the previous rating based upon the crude death rates. Minneapolis and Cleveland retain the first and the third place respectively. Omaha has dropped from the second to the fourth place, and Los Angeles jumped from the fifth to the second.

Infant Mortality. The number of children dying under one year of age is frequently used as an index of the health of a community. This is not a complex index, because it measures only a few of the factors or conditions which require consideration, and the same can be said of all the other indices which will be taken up in this paper. Still the attitude taken by a community toward the health and care of infants indicates in a certain way the general attitude toward all health activities. A study, therefore, of the infant mortality, should give some indication, even if in a general way, of the healthfulness of a given place. A serious difficulty,

Digitized by Google

however, is experienced in calculating a rate for expressing the infant mortality which will be reliable for purposes of comparison. It is fairly well agreed that the infant mortality should be expressed as so many deaths per 1,000 living births. But as the registration of births in the United States is notoriously defective, any rates based upon this registration will of course be inaccurate and not of much service. Then, too, it would not be fair to compare, on this basis, cities which have complete registration of births with those which have not. Since some of the cities studied do not have complete registration of births, unless perhaps for the most recent years, this method of expressing the infant mortality could not be used.

Several other methods were tried. The infant mortality was expressed in the following ways: (1) per 1,000 of living population under one year of age; (2) per 1,000 of living population under 5 years of age; (3) as a percentage of the total number of deaths; and (4) per 100,000 total population. It is obvious that all these are likely to be unsatisfactory. In the first two methods, the populations used in calculating the rates except for censal years must be estimated and these estimates are very likely to be inaccurate. Then, too, even in census years, the actual enumeration of persons in the age group under one is not considered to be very accurate. As the enumeration of persons in the age group 1–5 is likely to be more accurate, the second method was tried, but it did not prove to be of much value. The third method is even worse than the other two, for it is a ratio between two variable factors and so does not give any reliable picture of the infant mortality. If, for instance, the deaths from causes other than infant mortality decreased and the deaths of infants remained the same, the percentage obtained would indicate that infant mortality had increased, while in reality it would have remained the same. The fourth method is influenced by variations in the composition of the population and, therefore, is not entirely reliable.

The infant mortality was calculated, however, by all these methods to see what information could be gained. The accuracy of these methods was then tested by comparing the

Digitized by Google

results obtained by expressing the infant mortality as so many deaths per 1,000 living births. It is assumed, of course, that this latter method is the most accurate, and so should be used as a standard. The Bureau of the Census, in Bulletin 112, 1911, indicated that 8 of the cities, Los Angeles, St. Louis, New York, Boston, Omaha, Cincinnati, Pittsburgh, and Buffalo, are considered to have fairly complete registration of births, so that a rate of infant mortality based on births would be approximately correct for these cities. Such a rate was taken for a given year and these cities arranged according to the magnitude of their infant mortality, beginning with the city having the lowest. The cities were then similarly arranged for the same year by using rates calculated by each of the 4 other methods. None of the 4 methods placed the cities in the same relative position as regards their infant mortality as that in which they were placed by the standard method. The following table may help to make these points clearer:

TABLE III.

INFANT MORTALITY RATES, 1910.

City.	Infant Deaths per 1,000 Births.	Rank.	Infant Deaths per 1,000 Population under 5.	Rank.	Infant Deaths per 100,000 Total Population	Rank.	Infant Deaths per 1,000 Population under 1.	Rank.	Per Cent. of Infant Deaths to Total Deaths.	Rank.
Los Angeles.....	97	1	22.9	1	164	1	110	1	11.7	1
St. Louis.....	115	2	28.0	2	244	3	136	2	15.6	3
New York.....	125	3	31.8	5	337	6	147	4	21.1	6
Boston.....	126	4	35.1	6	333	5	167	6	19.5	5
Omaha.....	126	5	28.2	3	237	2	140	3	15.7	4
Cincinnati.....	131	6	31.4	4	252	4	150	5	14.5	2
Pittsburgh.....	150	7	38.9	8	421	8	176	7	23.4	7
Buffalo.....	162	8	38.2	7	382	7	180	8	23.4	8

It will be seen that the methods of expressing the infant mortality other than by the relation of infant deaths to births are not satisfactory for purpose of comparison between places. In view of this the rates as calculated by the 4 other methods will not be used in this paper. It should be added, however, that the calculation showed, for practically every city, by each of the 4 methods, that the infant mortality has decreased, and in some cases to a large extent, during the period 1900–1914.

Digitized by Google

Typhoid Fever. The death rate from typhoid fever is generally thought to be a good sanitary index of a community. It is a valuable index because it is a disease which can and should be controlled, for the principles underlying its eradication are fairly well understood and are possible of application. The data regarding typhoid fever are presented in the following table:

TABLE IV.

TYPHOID FEVER DEATH RATES PER 100,000 POPULATION.

City.	Average Rates, 1900–1914.	Average Rates, 1900–1902.	Average Rates, 1912–1914.	Decrease.
Baltimore	32.4	36.6	23.5	−13.1
Boston	16.7	23.6	8.4	−15.2
Buffalo	23.5	28.3	14.3	−14.0
Cincinnati	33.0	51.8	6.4	−45.4
Cleveland	28.1	41.9	9.8	−32.1
Denver	34.0	47.2	12.2	−35.0
Los Angeles	22.8	33.9	11.6	−22.3
Minneapolis	29.2	42.3	12.1	−30.2
New Orleans	34.4	48.9	17.5	−31.4
New York	14.2	20.4	7.6	−12.8
Omaha	24.8	23.3	9.0	−14.3
Pittsburgh	81.3	132.0	15.9	−116.1
San Francisco	23.8	28.1	14.4	−13.7
St. Louis	23.8	35.4	12.9	−22.5

Table IV plainly shows that for all the cities there has been a marked decrease in the amount of typhoid fever. Still the rates are much too high. Arranging the cities according to their typhoid rates for the period 1912–1914, we obtain the following relative standing:

(1) Cincinnati	6.4	(8) Denver	12.2
(2) New York	7.6	(9) St. Louis	12.9
(3) Boston	8.4	(10) Buffalo	14.3
(4) Omaha	9.0	(11) San Francisco	14.4
(5) Cleveland	9.8	(12) Pittsburgh	15.9
(6) Los Angeles	11.6	(13) New Orleans	17.5
(7) Minneapolis	12.1	(14) Baltimore	23.5

By this method of rating cities we should judge that Cincinnati was the most healthful, with New York second and Boston third. None of these cities, however, was in any of the first three positions when the rating was based upon the standardized death rates.

Tuberculosis. The death rate from tuberculosis is also a frequently used and valuable index in judging the sanitary

Digitized by Google

conditions of different places. The nature of the disease is fairly well understood. There are also certain procedures which are generally admitted to be efficacious in preventing the disease, and there are many organizations engaged in the work of prevention. The death rate from tuberculosis, therefore, is a valuable index of what a community is accomplishing along at least one line of health work as well as of its general healthfulness.

Only the pulmonary form of tuberculosis is considered, for this is the most common type of the disease and the one against which most of the preventive measures are directed. Pulmonary tuberculosis is also representative of the whole tuberculosis problem.

TABLE V.

PULMONARY TUBERCULOSIS DEATH RATES PER 100,000 POPULATION.

City.	Average Rates, 1900–1914.	Average Rates, 1900–1902.	Average Rates, 1912–1914.	Increase or Decrease.
Baltimore	223	232	191	− 41
Boston	185	232	146	− 86
Buffalo	125	119	130	+ 11
Cincinnati	235	214	220	+ 6
Cleveland	120	116	113	− 3
Denver	334	375	246	−129
Los Angeles	275	328	227	−101
Minneapolis	113	118	115	− 3
New Orleans	282	323	249	− 74
New York	199	226	169	− 57
Omaha	101	97	93	− 4
Pittsburgh	120	128	101	− 27
San Francisco	214	284	166	−118
St. Louis	175	187	136	− 51

In studying this table it must be remembered that in some instances the registration of deaths was incomplete for the period 1900–1902, and so the rates are too low. This tends to make the apparent decrease in the rates too small and may even account for an increase. A possible underestimate of the populations for 1912–1914 would have a similar effect upon the amount of decrease or increase. In general the rates have shown a tendency to decrease for most of the cities, and they probably have actually decreased in every instance. The cities have the following relative standing when arranged according to their pulmonary tuberculosis rates for the period 1912–1914:

Digitized by Google

(1) Omaha	93	(8) San Francisco	166
(2) Pittsburgh	101	(9) New York	169
(3) Cleveland	113	(10) Baltimore	191
(4) Minneapolis	115	(11) Cincinnati	220
(5) Buffalo	130	(12) Los Angeles	227
(6) St. Louis	136	(13) Denver	246
(7) Boston	146	(14) New Orleans	249

It will be observed that the positions of the cities have again changed to a marked extent, and do not correspond with any of the previous ratings.

Death Rates for the Age Group 1–4. There are various agencies engaged in the prevention of infant mortality. When a child goes to school its health is again cared for by the various health activities of the school organization. There is the possibility, however, that the health of children in the age group 1–4 is not sufficiently cared for, as there seems to be a gap here in the health activities for such children. Thus the death rates for this group age may serve as a valuable index of the inherent healthfulness of a given place. In any event it measures one of the factors which go to make up the healthfulness of a community.

TABLE VI.

DEATH RATES PER 100,000 POPULATION, IN AGE GROUP 1–4.

City.	Average Rates, 1900–1914.	Average Rates, 1900–1902.	Average Rates, 1912–1914.	Decrease.
Baltimore	156	185	122	− 63
Boston	153	213	111	−102
Buffalo	128	152	96	− 56
Cincinnati	123	161	81	− 80
Cleveland	129	155	114	− 41
Denver	84	117	52	− 65
Los Angeles	73	83	65	− 18
Minneapolis	73	93	64	− 29
New Orleans	136	174	104	− 74
New York	190	251	128	−123
Omaha	71	95	57	− 38
Pittsburgh	219	265	168	− 97
San Francisco	84	112	70	− 42
St. Louis	119	183	93	− 90

The question arose as to whether the rates should be expressed as so many deaths per 100,000 total population or per 1,000 under 5 years of age. The first method seemed preferable for the reasons previously noted. It must be remembered, however, that the age distribution of the population exerts

Digitized by Google

some influence on the rates when this method is followed, but the error from this cause is probably slight.

Table VI shows that for all the cities there has been a decrease in the death rates for the age group 1–4. Rating the cities according to this index, we obtain the following result:

(1) Denver	52	(8) Buffalo	96
(2) Omaha	57	(9) New Orleans	104
(3) Minneapolis	64	(10) Boston	111
(4) Los Angeles	65	(11) Cleveland	114
(5) San Francisco	70	(12) Baltimore	122
(6) Cincinnati	81	(13) New York	128
(7) St. Louis	93	(14) Pittsburgh	168

The relative positions of the cities are quite different from those shown in any of the other ratings.

Diarrhoea (Under 2). The death rate from diarrhoea under 2 years is frequently considered to be a good index of the sanitary conditions of a community, because it measures the care given children. The rates will be expressed as so many deaths per 100,000 total population. The following table presents the data relating to diarrhoea under two years of age:

TABLE VII.

DIARRHOEA (UNDER TWO) DEATH RATES PER 100,000 POPULATION.

City.	Average Rates, 1900–1914.	Average Rates, 1900–1902.	Average Rates, 1912–1914.	Increase or Decrease.
Baltimore	119	143	95	−48
Boston	100	119	74	−45
Buffalo	125	114	125	+11
Cincinnati	76	72	62	−10
Cleveland	120	95	131	+36
Denver	44	51	21	−30
Los Angeles	35	35	33	− 2
Minneapolis	47	51	36	−15
New Orleans	117	111	90	−21
New York	125	156	72	−84
Omaha	45	48	25	−23
Pittsburgh	158	171	118	−53
San Francisco	48	57	36	−21
St. Louis	68	59	63	+ 4

This method places the cities in the following relative positions:

(1) Denver	21	(3) Los Angeles	33
(2) Omaha	25	(4) San Francisco	36

Digitized by Google

(5) Minneapolis	36	(10) New Orleans	90
(6) Cincinnati	62	(11) Baltimore	95
(7) St. Louis	63	(12) Pittsburgh	118
(8) New York	72	(13) Buffalo	125
(9) Boston	74	(14) Cleveland	135

It will be noted that this rating is quite different from that for typhoid fever. From this it is probably fair to presume that the conditions affecting the prevalence of typhoid fever are not the same as those for diarrhoea.

The diseases which remain to be considered belong in a somewhat different class from those already covered. There are certain methods which are well understood and which are generally admitted by all to be efficacious in preventing typhoid fever and tuberculosis. There seems, however, to be no unanimity in regard to the best methods for restricting or eliminating the following diseases, with the possible exception of diphtheria, and the methods which are used do not give especially satisfactory results.

TABLE VIII.

MEASLES DEATH RATES PER 100,000 TOTAL POPULATION

City.	Average Rates, 1900–1914.	Average Rates, 1900–1902.	Average Rates, 1912–1914.	Increase or Decrease.
Baltimore	8.0	5.4	7.6	+ 2.2
Boston	15.0	17.2	11.6	− 5.6
Buffalo	12.5	13.1	11.7	− 1.4
Cincinnati	11.6	12.2	8.5	− 3.7
Cleveland	8.2	3.6	8.7	+ 5.1
Denver	7.7	8.8	1.2	− 7.6
Los Angeles	5.2	3.8	4.9	+ 1.1
Minneapolis	5.0	1.6	3.9	+ 2.3
New Orleans	6.8	7.3	4.3	− 3.0
New York	15.7	16.2	12.2	− 4.0
Omaha	5.7	5.0	.2	− 4.8
Pittsburgh	25.5	32.6	21.2	−11.4
San Francisco	8.2	9.9	7.6	− 2.3
St. Louis	9.3	2.5	9.3	+ 6.8

Table VIII shows that deaths from measles are a very variable factor in judging the health conditions of a place. In some instances the rates have increased during the last 15 years, and in others they have decreased. A study of the measles rates gives the following relative positions to the cities:

Digitized by Google

(1) Omaha	.2	(8) Cincinnati	8.5
(2) Denver	1.2	(9) Cleveland	8.7
(3) Minneapolis	3.9	(10) St. Louis	9.3
(4) New Orleans	4.3	(11) Boston	11.6
(5) Los Angeles	4.9	(12) Buffalo	11.7
(6) Baltimore	7.6	(13) New York	12.2
(7) San Francisco	7.6	(14) Pittsburgh	21.2

In a general way the higher rates are associated with the larger cities and the lower rates with the smaller cities, thus seeming to imply a relation between the density of population and the prevalence of measles.

TABLE IX.

SCARLET FEVER DEATH RATES PER 100,000 TOTAL POPULATION.

City.	Average Rates, 1900–1914.	Average Rates, 1900–1902.	Average Rates, 1912–1914.	Increase or Decrease.
Baltimore	8.4	4.4	5.6	+ 1.2
Boston	14.2	30.5	8.0	−22.5
Buffalo	13.7	7.7	4.8	− 2.9
Cincinnati	9.2	11.3	6.7	− 4.6
Cleveland	12.6	9.9	14.7	+ 4.8
Denver	19.3	29.4	13.2	−16.2
Los Angeles	4.4	6.6	2.1	− 4.5
Minneapolis	10.3	9.7	16.5	+ 6.8
New Orleans	5.5	10.8	1.2	− 9.6
New York	18.3	24.3	10.2	−14.1
Omaha	8.6	10.8	8.1	− 2.7
Pittsburgh	20.8	25.1	23.2	− 1.9
San Francisco	2.9	4.0	1.6	− 2.4
St. Louis	12.2	14.1	9.9	− 4.2

Table IX indicates that for most of the cities there has been a slight decrease in the death rates from scarlet fever. This would presumably indicate that there had been a decrease in the amount of scarlet fever. This, however, is not necessarily so, for the mortality from this disease varies in different epidemics and from year to year. The same is true of the other diseases in the group now being studied. It might appear to be better to use the number of cases of each disease in forming a judgment about the healthfulness of a place, but the fact that the number of cases which are reported for a disease is not nearly so complete as the number of deaths registered for the same disease is a serious objection to the use of this method. It may be noted in passing that while this paper is primarily concerned with judging healthfulness by a study of death rates, the author realizes that there may be other methods of attaining the same result.

Digitized by Google

A study of the scarlet fever rates places the cities in the following relative positions:

(1) New Orleans	1.2	(8) Omaha	8.1
(2) San Francisco	1.6	(9) St. Louis	9.9
(3) Los Angeles	2.1	(10) New York	10.2
(4) Buffalo	4.8	(11) Denver	13.2
(5) Baltimore	5.6	(12) Cleveland	14.7
(6) Cincinnati	6.7	(13) Minneapolis	16.5
(7) Boston	8.0	(14) Pittsburgh	23.2

It will be noticed that this rating is entirely different from that which obtains with measles.

TABLE X.

WHOOPING COUGH DEATH RATES PER 100,000 TOTAL POPULATION.

City.	Average Rates, 1900–1914.	Average Rates, 1900–1902.	Average Rates, 1912–1914.	Increase or Decrease.
Baltimore	12.4	13.5	10.1	− 3.4
Boston	14.5	18.0	10.1	− 7.9
Buffalo	10.5	10.7	9.7	− 1.0
Cincinnati	7.4	6.0	10.5	+ 4.5
Cleveland	7.9	5.8	7.6	+ 1.8
Denver	8.0	7.0	5.7	− 1.3
Los Angeles	6.6	9.3	6.1	− 3.2
Minneapolis	7.3	7.6	7.0	− 0.6
New Orleans	7.5	5.1	3.4	− 1.7
New York	8.0	12.2	6.2	− 6.0
Omaha	10.0	12.0	5.8	− 6.2
Pittsburgh	21.6	25.1	12.2	−12.9
San Francisco	9.8	8.1	7.3	− 0.8
St. Louis	7.4	6.5	7.4	+ 0.9

In most instances it will be seen that there has been a decrease in the death rates from whooping cough. The following is the list of cities arranged as to their average death rates from whooping cough for the period 1912–1914:

(1) New Orleans	3.4	(8) St. Louis	7.4
(2) Denver	5.7	(9) Cleveland	7.6
(3) Omaha	5.8	(10) Buffalo	9.7
(4) Los Angeles	6.1	(11) Baltimore	10.1
(5) New York	6.2	(12) Boston	10.1
(6) Minneapolis	7.0	(13) Cincinnati	10.5
(7) San Francisco	7.3	(14) Pittsburgh	12.2

The relative positions of the cities in this list are somewhat similar to those in the list based upon the measles death rates.

Digitized by Google

TABLE XI.

DIPHTHERIA DEATH RATES PER 100,000 TOTAL POPULATION.

City.	Average Rates, 1900–1914.	Average Rates, 1900–1902.	Average Rates, 1912–1914.	Increase or Decrease.
Baltimore	20.9	38.3	15.2	−23.1
Boston	34.0	63.6	20.0	−43.6
Buffalo	24.0	32.0	12.9	−19.1
Cincinnati	17.3	24.2	15.6	− 8.6
Cleveland	31.2	50.4	29.9	−20.5
Denver	21.3	33.4	4.8	−28.6
Los Angeles	16.7	26.7	7.5	−19.2
Minneapolis	33.2	61.0	24.3	−36.7
New Orleans	15.2	14.3	26.4	+12.1
New York	41.1	58.9	25.3	−33.6
Omaha	17.8	17.9	23.0	+ 5.1
Pittsburgh	34.3	46.6	32.1	−14.5
San Francisco	19.9	42.4	10.6	−31.8
St. Louis	28.4	47.9	27.8	−20.1

In all but two cases it will be seen that there has been a decline in the diphtheria death rates. This decrease has probably been in large measure due to the increased use of diphtheria antitoxin. By arranging the cities according to their diphtheria rates, the following result is obtained:

(1) Denver	4.8	(8) Omaha	23.0
(2) Los Angeles	7.5	(9) Minneapolis	24.3
(3) San Francisco	10.6	(10) New York	25.3
(4) Buffalo	12.9	(11) New Orleans	26.4
(5) Baltimore	15.2	(12) St. Louis	27.8
(6) Cincinnati	15.6	(13) Cleveland	29.9
(7) Boston	20.0	(14) Pittsburgh	32.1

This relative position of the cities somewhat resembles that obtained from the scarlet fever rates.

TABLE XII.

PNEUMONIA DEATH RATES PER 100,000 TOTAL POPULATION.

City.	Average Rates, 1900–1914.	Average Rates, 1900–1902.	Average Rates, 1912–1914.	Increase or Decrease.
Baltimore	196	207	194	−13
Boston	196	198	198	0
Buffalo	143	136	140	+ 4
Cincinnati	158	161	149	−12
Cleveland	133	166	114	−52
Denver	161	200	116	−84
Los Angeles	110	143	93	−50
Minneapolis	102	96	107	+11
New Orleans	170	185	156	−29
New York	237	276	190	−86
Omaha	122	122	131	+ 9
Pittsburgh	253	250	261	+11
San Francisco	154	181	131	−50
St. Louis	156	178	164	−14

Digitized by Google

Table XII plainly shows that in a general way the rates from pneumonia are slightly decreasing. The rates, however, clearly indicate that there is still much to be done in reducing the prevalence of this disease. The problem demands more attention from health authorities.

The pneumonia death rates give the following relative position to the cities:

(1) Los Angeles	93	(8) Cincinnati	149
(2) Minneapolis	107	(9) New Orleans	156
(3) Cleveland	114	(10) St. Louis	164
(4) Denver	116	(11) New York	190
(5) Omaha	131	(12) Baltimore	194
(6) San Francisco	131	(13) Boston	198
(7) Buffalo	140	(14) Pittsburgh	261

It will be readily observed that this rating is somewhat different from those obtaining for the other infectious diseases of the group being studied.

The results obtained by using the various methods are evidently not altogether comparable. That is, the cities do not consistently retain the same relative positions in the ratings. If a city is judged by one method it may hold first position; if judged by another method it may be third or fourth. All this goes to prove that any index based upon only one of the factors making up the health of a community is not an entirely satisfactory and just measure of its healthfulness.

The following table is presented to summarize the main points brought out in this paper. It brings together the relative positions obtained by each city from all the various rates studied. In addition, for purposes of comparison the relative position of each city is shown when an average of all its rates is taken as a basis for the grading.

Digitized by Google

TABLE XIII.

RELATIVE POSITIONS OF THE CITIES AS OBTAINED FROM VARIOUS DEATH RATES PREVIOUSLY PRESENTED.

City.	Typhoid.	Tuberculosis.	Age Group (1–4).	Diarrhœa (Under 2).	Measles.	Scarlet Fever.	Whooping Cough.	Diphtheria.	Pneumonia.	Total Average Rating.	Standardized Death Rates.
Baltimore....	14	10	12	11	6	5	11	5	12	13	13
Boston.....	3	7	10	9	11	7	12	7	13	10	9
Buffalo.....	10	5	8	13	12	4	10	4	7	8	8
Cincinnati...	1	11	6	6	8	6	13	6	8	9	11
Cleveland....	5	3	11	14	9	12	9	13	3	7	3
Denver.....	8	13	1	1	2	11	2	1	4	5	5
Los Angeles..	6	12	4	3	5	3	4	2	1	4	2
Minneapolis..	7	4	3	5	3	13	6	9	2	2	1
New Orleans.	13	14	9	10	4	1	1	11	9	12	14
New York ...	2	9	13	8	13	10	5	10	11	11	6
Omaha.....	4	1	2	2	1	8	3	8	5	1	4
Pittsburgh...	12	2	14	12	14	14	14	14	14	14	12
San Francisco	11	8	5	4	7	2	7	3	6	3	10
St. Louis....	9	6	7	7	10	9	8	12	10	6	7

If Table XIII is read from left to right, some very interesting information is obtained. It shows the good and the weak points in the healthfulness of each city. It shows very plainly that the health of a city may be very good when viewed from one angle, but very poor when viewed from another. Take the following instances: When typhoid fever and tuberculosis are considered, Cincinnati holds the first and the eleventh positions respectively. Pittsburgh, as a rule, has a poor rating, but that from tuberculosis is very good. In the same way New Orleans, which generally has a poor rating, stands very well when scarlet fever and whooping cough are considered.

Table XIII indicates in a way the line of health work which should be most energetically undertaken by each city. A city with a poor rating from typhoid fever needs to devote more attention and more work to the problem of eradicating this disease than to diseases which give it a good standing.

The tables also furnish a basis on which to select a living place. A person with young children would not select Cleveland as a place in which to live because of the poor standing which this city has in the matter of diarrhœa of early childhood. In the same manner, anyone who wished to escape tuberculosis would not select New Orleans, and so on.

By comparing the average ratings and those based upon

Digitized by Google

the standardized death rates, a general idea can be obtained of the relative ranking of the cities considered from all points of view.

In conclusion it seems reasonable to make the following deductions:

1. The use of a life table furnishes the best single means of comparing communities in regard to their health.

2. The use of standardized death rates constitutes the next best method, and it is much to be preferred to the use of the crude death rates.

3. The infant mortality rate is not a satisfactory index of the health of the community, because in most instances there is no accurate method of expressing it.

4. The death rates from the various infectious diseases do not give comparable results. The death rate from a single disease apparently does not give a just and fair representation of the healthfulness of a city. In judging the healthfulness of a city, such rates should therefore be used with care and discretion.

Digitized by Google

THE IMPROVEMENT OF STATISTICS OF CAUSE OF DEATH THROUGH SUPPLEMENTARY INQUIRIES TO PHYSICIANS.

By Louis I. Dublin, *Statistician*, and Edwin W. Kopf, *Chief Clerk, Statistical Bureau, Metropolitan Life Insurance Company, New York.*

The extension of the registration of vital statistics and the consequent increase in the amount of available data on American mortality have resulted in recent years in the development of a critical spirit toward these statistics. When data were scant, they were perforce accepted with little or no question; today we are justified in inquiring to what extent the returns are an accurate picture of the actual conditions. Perhaps enough material is now at hand to enable one to separate the probably true from the uncertain in our mortality returns. This questioning is clearly warranted by the likelihood that the sources of our statistics will thus be improved. It is only as our basic material is sound that the death rates we quote have real value.

The chief source of error in the statistics of mortality is, of course, the uncertainty of the primary diagnosis made by the physician certifying to the cause of death. A group of clinicians and pathologists, represented by Cabot,* Oertel,† and Emerson,‡ have in recent years contributed stimulating and constructive studies of the general reliability of diagnoses made in fatal cases. These writers point out, for example, that the acute infections, like typhoid fever, scarlet fever, and diphtheria, and those conditions which are usually of long duration and present distinct clinical pictures, like pulmonary tuberculosis and the external cancers, furnish no special difficulties of diagnosis to the physician, and statements with regard to them may be accepted without further question. On the other hand, bedside diagnoses, even in our best hospi-

* Cabot, Richard C., M.D.: Diagnostic Pitfalls I dentified during a Study of Three Thousand Autopsies Journal Amer. Medical Ass'n, December 28, 1912.

† Oertel, Horst, M.D.: The Inaccuracy of American Mortality Statistics, American Underwriter, May, 1913.

‡ Emerson, Haven, M.D.: Reliability of Certain Classes of Death Certificates, Amer. Jour. Public Health, August, 1915.

Digitized by Google

tals, of Bright's disease, of cerebral hemorrhage and apoplexy, and of certain of the heart affections are very frequently not confirmed by autopsy, while other determinative conditions are shown to have been the primary cause of death. Relief from this fundamental difficulty is not within the control of statisticians, but will depend entirely upon the improvement of autopsy facilities and the general rise of standards of medical education. A Committee of the Vital Statistics Section of the American Public Health Association is now actively engaged in considering the International List of Causes of Death from this point of view, registering its conclusions as to those titles which can be accepted with considerable certainty even in the absence of an autopsy, as against those which must be considered as tentative assignments only, unless based on autopsy findings.

The present paper bears on the remedy for a second important source of error which is largely within the scope and control of statisticians. Fortunately, this remedy can be applied at once and surprisingly large corrections in the death rate for certain causes can be made available. The method of procedure consists in attempting to secure from physicians completer statements as to the cause of death, in those cases where the returns on the death certificate are given vaguely or indefinitely, or where there is evidence that the original statements do not tell the entire story. Thus physicians often report conditions like "acute nephritis" as the cause of death, and do not refer at all to the primary condition, such as "scarlet fever," "influenza," or "alcoholism," which was present and known to them. Again, "peritonitis" is frequently reported without reference to the "appendicitis" or to the "puerperal septicemia" which may have preceded it. The various forms of death by violence are still reported vaguely, and without due regard for the statistical requirements of assigment to the International list of Causes of Death. Certificates frequently do not permit one to determine whether death was due to accident, suicide, or homicide, and what was the means and character of the injury.

This report is not the announcement of a discovery in the field of statistics. It is simply an attempt to put on record

Digitized by Google

the results of a method followed for four years in the Statistical Bureau of the Metropolitan Life Insurance Company, in the treatment of the industrial mortality experience of that company. We have realized that, if our material on the mortality of the American working classes is to be utilized to the full, every effort must be made to improve the quality of the basic data, especially since for many diseases and conditions the death rate is decidedly modified by the extent to which the query system is applied. The method is essentially that followed by the Registrar-General of England and Wales, and by the Division of Vital Statistics in the United States Census Bureau. We desire at this point to acknowledge our indebtedness to Mr. George H. Van Buren, Chief of that Division, who aided materially in the introduction of the method into our office. Both the English* and the American† bureaus have already, in their annuals and in other publications, called attention to the changes resulting from their systems of inquiry. A preliminary statement of the results obtained by our office appeared in the December, 1913, issue of this QUARTERLY.‡ In the present paper we shall discuss the effects upon our returns on the basis of the much more extensive data now available. It is our hope that this complete statement will encourage other workers to similar efforts, and that in this way the returns of health departments and insurance companies may become more accurate and comparable.

The technical procedure is as follows: The death certificates attached to the claim papers are carefully examined, and the staff engaged in this work determine whether the statement of cause of death given by the physician is satisfactory and complete for purposes of assignment. Statements of cause presenting no difficulty are at once classified according to the International List of Causes of Death. Joint causes are classified according to the precedents and rules of the International Committee, as best exemplified by the methods of the United

* Registrar-General of Births, Deaths and Marriages. Report, 1913, p. LXXIX.

† Bureau of the Census. Mortality Statistics, 1911, p. 36.
Bureau of the Census. Mortality Statistics, 1912, p. 23.
Van Buren, Geo. H.: Mortality Statistics and the Physician, Amer. Jour. Pub. Health, Dec., 1915.
Emerson, Haven, *loc. cit.*

‡ Dublin, Louis I.; Kopf, Edwin W.: Experiment in the Compilation of Mortality Statistics, this Quarterly, Dec., 1913.

Digitized by Google

States Census Bureau. Wherever the statement of cause is considered unsatisfactory for classification, or where, from the statement presented to the Company by the claimant or from other available sources, there is an indication that a serious condition not mentioned on the certificates has contributed to the death, or where additional information is required before a satisfactory assignment can be made, a letter of inquiry, asking for the facts, is sent to the physician. Clerical labor is minimized wherever possible through the use of form letters. A codified series of questions is employed, and the examiners indicate to the typists the question which is to be inserted in the form. The following is a copy of one of our series of letters sent to physicians:

NEW YORK CITY

Dear Doctor:

This Company is analyzing its mortality experience for the year 191 , and in order that we may be able to classify accurately the death of who died on , and whose certificate you signed, we most respectfully ask you to send us the following information in addition to the statement now appearing on the death claim, viz.:

Acute nephritis

Additional data desired:

Kindly indicate the primary cause of the ACUTE NEPHRITIS. Due to any contagious or infectious disease (scarlet fever, typhoid fever, etc.)? *Directly* due to alcoholism or exposure? Traumatic? If so, means of injury?

The claim was duly paid by this Company under date of 191 , and this desired information will be used for statistical purposes only. We are confident that your interest in the advancement of the scientific study of Causes of Death will prompt you to comply with our request, and we beg you to accept our thanks. A stamped envelope is enclosed for your reply.

Very respectfully yours,

Statistician.

(Please answer here)
15

At first we proceeded cautiously, inquiring only as to the means and character of the injury in cases of ill-defined violence. We also asked for the causes for the relief of which operations were performed (when expressions such as "operation," "surgical shock," or "laparotomy" were reported with no additional data), the causes contributing to "peritonitis," and the

Digitized by Google

more exact location of cancers and tumors. We asked for additional information when the terms were "dropsy," "heart failure," and others of like character included in the designation "Ill-defined Diseases" in the International Classification. The scope of our questions has since been much extended. The following questions indicate the lines along which inquiry is now made:

UNDESIRABLE RETURNS AND CHARACTER OF INQUIRY MADE.

Undesirable Returns.	*Inquiry.*
1. "Injury," "Gunshot-Wound," "Drowning," "Accidental death," or other ill-defined violence.	Was death due to accident, suicide, or homicide? If accident, what kind of accident was the primary cause of death (steam railroad, street car, fall, etc.)?
2. "Operation," "Hysterectomy," "Laparatomy."	What was the cause, for the relief of which the operation was performed?
3. "Peritonitis."	Was the peritonitis secondary to typhoid fever? To tuberculosis? To puerperal septicemia? To appendicitis? If postoperative, what was the primary cause? Was it accidental, suicidal, or homicidal? If so, what was the means of injury?
4 "Acute nephritis."	What was the primary cause? Was it a contagious or infectious disease (scarlet fever, typhoid fever, etc.)? Was it directly due to alcoholism or exposure? Was it traumatic? If so, what was the means of injury?
5. "Meningitis," or "Cerebrospinal meningitis."	Was it epidemic? Tuberculous? Was it simple meningitis following lobar pneumonia, typhoid fever, or any other contagious or infectious disease? Was it traumatic? If so, what were the means and nature of the injury? Was it accidental, suicidal, or homicidal?
6. "Apoplexy."	Was this a case of cerebral hemorrhage?
7. "Paralysis."	Was it superinduced by cerebral hemorrhage? Was it a case of spinal paralysis? Of general paralysis of the insane?
8. "General paralysis of the insane."	Was it due to syphilis? If so, was the presence of syphilis determined by history or tests?
9. "Locomotor ataxia."	Was it of syphilitic origin? Traumatic? If so, what was the means of injury?

Digitized by Google

10. "Progressive paralysis."	Was it due to a spinal lesion? Was it directly or remotely due to cerebral hemorrhage?
11. "Tumor."	Was it malignant? What was its location?
12. "Cancer."	What was its location?
13. "Abscess."	What was its location? What was its nature and cause? Was it tuberculous? Traumatic?
14. "Burns."	Was death primarily due to burns received in burning building? (Such deaths are separately classified.)
15. Any terminal condition, origin not stated, such as "Septicemia," "Convulsions," or "Hemorrhage."	What was the primary cause?
16. "Acute cardiac dilatation," "Heart failure."	What was the primary cause?
17. "Pneumonia."	Was it lobar, bronchial, grippal, or tuberculous? Was it traumatic? If so, was death due to accident, suicide, or homicide, and what was the means of injury?

At the present time we are querying about 8.5 per cent. of the certificates of death received in our office. Satisfactory replies are obtained from physicians in about three quarters of the cases; the proportion varies somewhat with the condition queried. Changes are made in the classification, as the result of inquiry, in about 70 per cent. of the replies received. It is a pleasure at this time to acknowledge our indebtedness to the thousands of physicians whom we have addressed, and who, without reimbursement, have put at our disposal their time and opinion in order to help in the campaign to improve vital statistics. It is an encouraging commentary on the scientific interest of the American practitioner that such a large return could be received.

The following table shows the principal titles in which changes in assignment were made; it shows the number of deaths that would have been assigned to each title without inquiry, the number actually assigned after inquiry, and the ratio of the latter number to the former. The titles are arranged in the increasing order of these ratios.

Digitized by Google

TABLE I.

NUMBER OF DEATHS FROM CERTAIN CAUSES, BEFORE AND AFTER INQUIRY, AND PROPORTION OF LATTER TO FORMER ARRANGED IN INCREASING ORDER OF PROPORTIONS.

Metropolitan Life Insurance Company, Industrial Department; Mortality Experience, 1911–1914.

Cause of Death.		Deaths before Inquiry.	Deaths after Inquiry.	Deaths after Inquiry, per 100 Deaths before Inquiry.
International List Number.	Title.			
185	Fractures (causes not specified)	1,484	545	36.7
117	Simple peritonitis—nonpuerperal	1,180	711	60.3
066	Paralysis without specified cause	3,340	2,090	62.6
061	Simple meningitis	3,752	2,534	67.5
046	Other tumors (tumors of the female genital organs excepted)	266	196	73.7
062	Locomotor ataxia	714	558	78.2
186	Other external violence	1,456	1,158	79.5
067	General paralysis of the insane	2,042	1,627	79.7
119	Acute nephritis	4,186	3,375	80.6
020	Purulent infection and septicemia	820	725	88.4
094	Pulmonary congestion, pulmonary apoplexy	749	664	88.7
045	Cancer and other malignant tumors of other organs or of organs not specified	3,650	3,279	89.8
092	Pneumonia (lobar and undefined)	27,740	26,299	94.8
081	Diseases of the arteries, atheroma, aneurysm, etc.	6,390	6,199	97.0
120	Bright's disease	32,854	32,745	99.7
028	Tuberculosis of the lungs	60,486	60,613	100.2
009	Diphtheria and croup	8,850	8,881	100.4
156–163	Other suicides	2,660	2,674	100.5
029, 032–035	Other forms of tuberculosis	6,083	6,120	100.6
040	Cancer and other malignant tumors of the stomach	8,804	8,932	101.5
008	Whooping cough	1,995	2,025	101.5
001	Typhoid fever	6,343	6,440	101.5
155	Suicide by poison	1,666	1,696	101.8
006	Measles	3,137	3,216	102.5
007	Scarlet fever	3,681	3,778	102.6
039	Cancer and other malignant tumors of the buccal cavity	853	880	103.2
031	Abdominal tuberculosis	2,067	2,133	103.2
041	Cancer and other malignant tumors of the peritoneum, intestines, rectum	2,679	2,769	103.4
042	Cancer and other malignant tumors of the female genital organs	4,813	5,003	103.9
108	Appendicitis and typhlitis	3,807	3,959	104.0
043	Cancer and other malignant tumors of the breast	2,089	2,195	105.1
044	Cancer and other malignant tumors of the skin	574	604	105.2
064	Cerebral hemorrhage and apoplexy	21,727	23,029	106.0
182	Homicide by firearms	1,443	1,532	106.2
091	Bronchopneumonia	9,133	9,708	106.3
183–184	Homicide by other means	792	848	107.1
010	Influenza	3,995	4,370	109.4
A63	Other diseases of the spinal cord	1,248	1,389	111.3
030	Tuberculous meningitis	2,656	3,024	113.9
051	Exophthalmic goitre	368	433	117.7
056	Alcoholism (acute or chronic)	1,353	1,633	120.7
172	Traumatism by fall	3,575	4,432	124.0
037	Syphilis	1,322	2,350	177.8
038	Gonococcus infection	50	115	230.0
A61	Cerebrospinal fever	138	536	388.4

The indefinite title "fractures (cause not specified)" shows the greatest reduction in the number of assignments. The number of deaths after inquiry is only 36.7 per cent. of the

Digitized by Google

number originally returned. In view of the fact that the number of persons exposed is constant, the decrease in the number of deaths is accompanied by a corresponding reduction in the death rate. We may, therefore, say that the final death rate for this cause was 63.3 per cent. less than the one which would originally have been quoted. Such general titles as "other external violence" and "cancer of organs not specified" were in like manner greatly reduced through the determination of the nature of the violence, or of the location of the cancer. Considerable reductions were obtained upon inquiry into the character of such returns as "paralysis" and "locomotor ataxia"; for in such cases added information with reference to blood tests, etc., gave sufficient data upon which more definite assignments could be made. On the other hand, a large number of causes, especially those referring to the infectious diseases, show very little change. Tuberculosis of the lungs, for example, shows an increase of but .2 per cent. The infectious diseases of children, such as "measles" and "scarlet fever," give somewhat greater percentages of increase, 2.5 and 2.6 respectively; these increases were made at the expense of terminal conditions, such as "hypostatic pneumonia" and "acute nephritis," which are very often given as the causes of death. Cancers of definitely specified organs were considerably increased in the number of assignments, with a corresponding increase in the rates. Thus "cancer and other malignant tumors of the breast," of which there are many cases, showed an increase of 5.1 per cent. in the death rate. Such socially significant titles as "alcoholism," "syphilis," and "gonococcus infection" gave very marked rates of increase, 20.7, 77.8, and 130.0 per cent., respectively. That we could add two thirds as many deaths as were originally assigned to "syphilis" is in itself sufficient justification for the effort involved by our method. The greatest percentage of increase is shown by the title "cerebrospinal fever," which was increased from 138 to 536 deaths, or nearly fourfold.

The more definite titles to which some of the undesirable titles were changed are shown in the following table:

Digitized by Google

TABLE II.

ILL-DEFINED CAUSES OF DEATH, AND DEFINITE TITLES TO WHICH RE-ASSIGNED, BY NUMBER AND PERCENTAGE OF CHANGE IN CLASSIFICATION.

Metropolitan Life Insurance Company, Industrial Department; Mortality Experience, 1911-1914.

Ill-Defined Term as Classified before Inquiry.	Title Assigned after Inquiry.	Number and Percentage of Changes in Classification.	
		Number.	Percentage of Total Replies.
Purulent infection and septicemia (Total replies, 184.)	Total changes from "Puerperal infection and septicemia"	134	72.8
	"Other diseases of the uterus"	6	3.3
	Puerperal septicemia	54	29.3
	Acute abscess	8	4.3
	Other titles	66	35.9
Cancer and other malignant tumors of other organs or of organs not specified (Total replies, 572.)	Total changes from "Cancer . . . not specified"	423	74.0
	Cancer and other malignant tumors of the buccal cavity	25	4.4
	Cancer and other malignant tumors of the stomach	75	13.1
	Cancer and other malignant tumors of the peritoneum	61	10.7
	Cancer and other malignant tumors of the female genital organs	148	25.9
	Cancer and other malignant tumors of the breast	83	14.5
	Cancer and other malignant tumors of the skin	28	4.9
	Other titles	3	.5
Meningitis (Total replies, 1,653.)	Total changes from "Meningitis"	1,234	74.7
	Tuberculous meningitis	294	17.8
	Cerebrospinal fever	363	22.0
	Pneumonia	79	4.8
	Diarrhea and enteritis—2 years and over	62	3.8
	Other titles	436	26.4
Locomotor ataxia (Total replies, 321.)	Total changes from "Locomotor ataxia"	163	50.8
	Syphilis	148	46.1
	Other titles	15	4.7
Paralysis without specified cause (Total replies, 1,449.)	Total changes from "Paralysis without specified cause"	1,259	86.9
	"Other diseases of the spinal cord"	175	12.1
	Cerebral hemorrhage, apoplexy	885	61.1
	General paralysis of the insane	80	5.5
	Other titles	119	8.2
General paralysis of the insane (Total replies, 956.)	Total changes from "General paralysis of the insane"	531	55.5
	Syphilis	491	51.4
	Other titles	40	4.2
Organic diseases of the heart* (Total replies, 714.)	Total changes from "Organic diseases of the heart"	264	37.0
	Cerebral hemorrhage, apoplexy	46	6.4
	Pneumonia	72	10.1
	Other titles	146	20.4
Diseases of the arteries, atheroma, aneurysm, etc. (Total replies, 603.)	Total changes from "Diseases of arteries, etc."	317	52.6
	Syphilis	65	10.8
	Cerebral hemorrhage, apoplexy	166	27.5
	Other titles	86	14.3
Pneumonia† (Total replies, 3,005.)	Total changes from "Pneumonia"	1,754	58.4
	Influenza	228	7.6
	Bronchopneumonia	586	19.5
	Lobar pneumonia	577	19.2
	Other titles	363	12.1

* Most of the terms included under the title "organic diseases of the heart" are fairly definite. However, certain expressions like "acute cardiac dilatation" and "cardiac insufficiency" are often reported when they are only terminal symptoms of other diseases. It is these expressions that are covered by our inquiry.

†Assignments to "lobar pneumonia" should be regarded as confirmations of original assignments to "pneumonia," rather than as changes.

Digitized by Google

TABLE II—*Continued.*

Ill-Defined Term as Classified before Inquiry.	Title Assigned after Inquiry.	Number and Percentage of Changes in Classification.	
		Number.	Percentage of Total Replies.
Pulmonary congestion, pulmonary apoplexy (Total replies, 155.)	Total changes from "pulmonary apoplexy"	99	63.9
	Organic diseases of the heart	15	9.7
	Chronic bronchitis	6	3.9
	Pneumonia	8	5.2
	Bright's disease	8	5.2
	Other titles	62	40.0
Peritonitis — non-puerperal (Total replies, 575.)	Total changes from "Simple peritonitis"	476	82.8
	Abdominal tuberculosis	44	7.7
	Diarrhea and enteritis—2 years and over	20	3.5
	Appendicitis and typhlitis	105	18.3
	Salpingitis and other diseases of the female genital organs	45	7.8
	Puerperal septicemia	86	15.0
	Other titles	176	30.6
Acute nephritis (Total replies, 1,935.)	Total changes from "Acute nephritis"	839	43.4
	Scarlet fever	68	3.5
	Influenza	75	3.9
	Alcoholism (acute or chronic)	213	11.0
	Bright's disease	74	3.8
	Other titles	409	21.1
Other diseases of the uterus (Total replies, 173.)	Total changes from "Other diseases of the uterus"	119	68.8
	Gonococcus infection	15	8.7
	Cancer and other malignant tumors of the female genital organs	7	4.0
	Uterine tumor (noncancerous)	13	7.5
	Cysts and other tumors of the ovary	6	3.5
	Salpingitis and other diseases of the female genital organs	21	12.1
	Puerperal septicemia	35	20.2
	Other titles	22	12.7
Traumatism by firearms (Total replies, 119.)	Total changes from traumatism by firearms	95	79.8
	Suicide by firearms	6	5.0
	Homicide by firearms	87	73.1
	Other titles	2	1.7
Fractures (cause not specified) (Total replies, 1,045.)	Total changes from "Fractures (cause not specified)"	949	90.8
	Traumatism by fall	646	61.8
	Traumatism by crushing—electric railway	33	3.2
	Traumatism by crushing—steam railway	25	2.4
	Traumatism by crushing—automobiles	27	2.6
	Traumatism by crushing—other vehicles	51	4.9
	Traumatism by other means	57	5.5
	Other titles	110	10.5
Other external violence (Total replies, 504.)	Total changes from "Other external violence"	400	79.4
	Purulent infection and septicemia	15	3.0
	Traumatism by fall	161	31.9
	Traumatism by machines	23	4.6
	Traumatism by crushing—steam railway	29	5.8
	Traumatism by crushing—electric railway	16	3.2
	Traumatism by crushing—automobiles	17	3.4
	Traumatism by crushing—other vehicles	37	7.3
	Other titles	102	20.2
Ill-defined diseases (Total replies, 442.)	Total changes from "Ill-defined diseases"	322	72.9
	Malaria	15	3.4
	Tuberculosis of the lungs	18	4.1
	Organic diseases of the heart	83	18.8
	Bright's disease	21	4.8
	Other titles	185	41.9

Digitized by Google

Letters were received from physicians in connection with 1,653 cases originally diagnosed as "meningitis." In 1,234 of these cases, or 74.7 per cent., the classification was changed to other and more definite titles. "Cerebrospinal fever" received 22 per cent. of the original "meningitis" returns; "tuberculous meningitis," 17.8 per cent.; and "pneumonia (lobar and undefined)," 4.8 per cent. The title "paralysis without specified cause" was the subject of 1,449 replies; the classification was changed in 1,259 instances, or 86.9 per cent. The larger proportion (61.1 per cent.) of these "paralysis without specified cause" cases were classified under "cerebral hemorrhage and apoplexy." "Peritonitis" was changed in classification in 476 out of 575 cases, or 82.8 per cent. These changes were credited chiefly to the titles "appendicitis and typhlitis," "puerperal septicemia," "salpingitis," and other diseases of the female genital organs.

Ill-defined returns of "traumatism by firearms" were changed in 79.8 per cent. of the cases queried. The greater part (73.1 per cent.) of these deaths was assigned to the title "homicide by firearms." "Fractures (cause not specified)" were a fruitful source of additions to the class of "traumatism by fall"; 646 ill-defined fractures (61.8 per cent.) were so assigned.

After this examination of the ill-defined terms and the more definite titles to which they were assigned, we may proceed naturally to the other aspect of the change; namely, to determine how the mortality is constituted in those causes which show an increase in the number of deaths. This is indicated by Table III.

Digitized by Google

TABLE III.

ADDITIONS TO CERTAIN CAUSES OF DEATH FROM INDEFINITE TITLES, NUMBER AND PERCENTAGE CHANGED FROM EACH INDEFINITE TITLE.

Metropolitan Life Insurance Company, Industrial Department; Mortality Experience, 1911–1914.

Cause of Death Title Receiving Additions.	Title before Inquiry.	Additions.	
		Number.	Per Cent.
Typhoid fever	Total changes to "typhoid fever"	161	100.0
	Simple meningitis	29	28.7
	Organic diseases of the heart	3	3.0
	Pneumonia	5	5.0
	Other diseases of the intestines	6	5.9
	Simple peritonitis	16	15.8
	Acute nephritis	21	20.8
	Bright's disease	3	3.0
	Cause of death not specified or ill-defined	3	3.0
	Other titles	15	14.9
Measles	Total changes to "measles"	89	100.0
	Whooping cough	3	3.4
	Simple meningitis	19	21.3
	Bronchopneumonia	9	10.1
	Pneumonia	31	34.8
	Acute nephritis	13	14.6
	Other titles	14	15.7
Scarlet fever	Total changes to "scarlet fever"	98	100.0
	Simple meningitis	7	7.1
	Bronchopneumonia	4	4.1
	Pneumonia	3	3.1
	Acute nephritis	68	69.4
	Bright's disease	5	5.1
	Other titles	11	11.2
Whooping cough	Total changes to "whooping cough"	45	100.0
	Simple meningitis	12	26.7
	Convulsions of infants	2	4.4
	Bronchopneumonia	6	13.3
	Pneumonia	14	31.1
	Pulmonary congestion, pulmonary apoplexy	2	4.4
	Acute nephritis	5	11.1
	Other titles	4	8.9
Diphtheria and croup	Total changes to "diphtheria and croup"	35	100.0
	Acute endocarditis	2	5.7
	Organic diseases of the heart	2	5.7
	Diseases of the larynx	6	17.1
	Acute bronchitis	2	5.7
	Bronchopneumonia	4	11.4
	Pneumonia	3	8.6
	Diseases of the pharynx	2	5.7
	Acute nephritis	10	28.6
	Other titles	4	11.4
Influenza	Total changes to "influenza"	394	100.0
	Simple meningitis	40	10.2
	Pneumonia	228	57.9
	Acute nephritis	75	19.0
	Other titles	51	12.9
Tuberculosis of the lungs	Total changes to "tuberculosis of the lungs"	166	100.0
	Simple meningitis	6	3.6
	Organic diseases of the heart	6	3.6
	Chronic bronchitis	6	3.6
	Pneumonia	70	42.2
	Cause of death not specified or ill-defined	16	9.6
	Other titles	62	37.3
Tuberculous meningitis	Total changes to "tuberculous meningitis"	377	100.0
	Meningitis	294	78.0
	Other titles	83	22.0

Digitized by Google

TABLE III—*Continued.*

Cause of Death Title Receiving Additions.	Title before Inquiry.	Additions. Number.	Additions. Per Cent.
Syphilis	Total changes to "syphilis"	**1,031**	**100.0**
	Locomotor ataxia	148	14.4
	Other diseases of the spinal cord	59	5.7
	Cerebral hemorrhage, apoplexy	34	3.3
	Paralysis without specified cause	34	3.3
	General paralysis of the insane	491	47.6
	Diseases of the nervous system other than those specified	96	9.3
	Diseases of the arteries, atheroma and aneurysm	65	6.3
	Other titles	104	10.1
Cancer and other malignant tumors of the stomach and liver	Total changes to "cancer and other malignant tumors of the stomach and liver"	**130**	**100.0**
	Cancer and other malignant tumors of other organs or of organs not specified	75	57.7
	Other tumors (tumors of the female genital organs excepted)	7	5.4
	Diseases of the esophagus	4	3.1
	Other diseases of the stomach	7	5.4
	Other diseases of the liver	13	10.0
	Other titles	24	18.5
Cancer and other malignant tumors of the female genital organs	Total changes to "cancer and other malignant tumors of the female genital organs"	**190**	**100.0**
	Cancer and other malignant tumors of other organs or of organs not specified	148	77.9
	Other tumors (tumors of the female genital organs excepted)	7	3.7
	Uterine tumor (noncancerous)	7	3.7
	Other diseases of the uterus	7	3.7
	Other titles	21	11.1
Cancer and other malignant tumors of the breast	Total changes to "cancer and other malignant tumors of the breast"	**106**	**100.0**
	Cancer and other malignant tumors of other organs or of organs not specified	83	78.3
	Nonpuerperal diseases of the breast (cancer excepted)	9	8.5
	Other titles	14	13.2
Alcoholism	Total changes to "alcoholism"	**289**	**100.0**
	Simple meningitis	14	4.8
	Acute nephritis	213	73.7
	Cause of death not specified or ill-defined	10	3.5
	Other titles	52	18.0
Cerebrospinal fever	Total changes to "cerebrospinal fever"	**401**	**100.0**
	Simple meningitis	363	90.5
	Other titles	38	9.5
Cerebral hemorrhage, apoplexy	Total changes to "cerebral hemorrhage, apoplexy"	**1,408**	**100.0**
	Paralysis without specified cause	885	62.9
	Organic diseases of the heart	46	3.3
	Diseases of the arteries, atheroma, aneurysm	166	11.8
	Bright's disease	137	9.7
	Other titles	174	12.4
Organic diseases of the heart*	Total changes to "organic diseases of the heart"	**287**	**100.0**
	Cerebral hemorrhage, apoplexy	10	3.5
	Paralysis without specified cause	13	4.5
	Acute endocarditis	10	3.5
	Diseases of the arteries, atheroma, aneurysm	14	4.9
	Pneumonia	12	4.2
	Pulmonary congestion, pulmonary apoplexy	15	5.2
	Acute nephritis	25	8.7
	Bright's disease	9	3.1
	Senility	13	4.5
	Fractures (cause not specified)	11	3.8
	Ill-defined organic disease	39	13.6
	Cause of death not specified or ill-defined	43	15.0
	Other titles	73	25.4

*See footnote on this title in Table II.

Digitized by Google

TABLE III—*Continued.*

Cause of Death Title Receiving Additions.	Title before Inquiry.	Additions.	
		Number.	Per Cent.
Diseases of the arteries, atheroma, aneurysm	Total changes to "diseases of the arteries, atheroma, aneurysm"	126	100.0
	Other diseases of the spinal cord	5	4.0
	Softening of the brain	9	7.1
	Paralysis without specified cause	9	7.1
	Organic diseases of the heart	6	4.8
	Pneumonia	5	4.0
	Pulmonary congestion, pulmonary apoplexy	4	3.2
	Acute nephritis	10	7.9
	Gangrene	28	22.2
	Senility	10	7.9
	Cause of death not specified or ill-defined	4	3.2
	Other titles	36	28.6
Bronchopneumonia	Total changes to "bronchopneumonia"	669	100.0
	Simple meningitis	32	4.8
	Pneumonia	586	88.8
	Other titles	42	6.4
Pneumonia (lobar and undefined)	Total changes to "pneumonia"	329	100.0
	Simple meningitis	79	24.0
	Organic diseases of the heart	72	21.9
	Pleurisy	13	4.0
	Simple peritonitis (nonpuerperal)	10	3.0
	Cerebral hemorrhage, apoplexy	10	3.0
	Bright's disease	44	13.4
	Other titles	101	30.7
Appendicitis and typhlitis	Total changes to "appendicitis and typhlitis"	156	100.0
	Intestinal obstruction	6	3.8
	Simple peritonitis (nonpuerperal)	105	67.3
	Other titles	45	28.8
Bright's disease	Total changes to "Bright's disease"	235	100.0
	Cerebral hemorrhage, apoplexy	7	3.0
	Organic diseases of the heart	19	8.1
	Diseases of the arteries, atheroma, aneurysm	13	5.5
	Pulmonary congestion, pulmonary apoplexy	8	3.4
	Acute nephritis	74	31.5
	Other diseases of the kidneys and annexa	12	5.1
	Senility	7	3.0
	Fractures (cause not specified)	9	3.8
	Ill-defined organic disease	16	6.8
	Other titles	70	29.8
Puerperal state—total	Total changes to "puerperal state—total"	368	100.0
	Purulent infection and septicemia	54	14.7
	Convulsions (nonpuerperal)	32	8.7
	Pneumonia	15	4.1
	Simple peritonitis (nonpuerperal)	87	23.6
	Acute nephritis	58	15.8
	Bright's disease	26	7.1
	Uterine hemorrhage (nonpuerperal)	4	1.1
	Other diseases of the uterus	40	10.9
	Other titles	52	14.1
Puerperal septicemia	Total changes to "puerperal septicemia"	189	100.0
	Purulent infection and septicemia	50	26.5
	Simple peritonitis	77	40.7
	Diseases of the uterus	32	16.9
	Other titles	30	15.9
Puerperal albuminuria and convulsions	Total changes to "puerperal albuminuria and convulsions"	125	100.0
	Convulsions (nonpuerperal)	32	25.6
	Acute nephritis	54	43.2
	Bright's disease	26	20.8
	Other titles	13	10.4

Digitized by Google

TABLE III—*Concluded.*

Cause of Death Title Receiving Additions.	Title before Inquiry.	Additions.	
		Number.	Per Cent.
Suicide	Total changes to "suicide"	46	100.0
	Acute poisonings	24	52.2
	Absorption of deleterious gases (conflagration excepted)	10	21.7
	Traumatism by firearms	6	13.0
	Fractures (cause not specified)	2	4.3
	Other titles	4	8.7
Traumatism by falling	Total changes to "traumatism by falling"	945	100.0
	Simple meningitis	43	4.6
	Fractures (cause not specified)	646	68.4
	Other external violence	161	17.0
	Other titles	95	10.1
Homicide—total	Total changes to "homicide"	148	100.0
	Simple peritonitis (nonpuerperal)	7	4.7
	Traumatism by firearms	87	58.8
	Fractures (cause not specified)	15	10.1
	Other external violence	16	10.8
	Other titles	23	15.5

Of the 101 cases added to typhoid fever, 28.7 per cent. were originally classified as "simple meningitis," 20.8 per cent. as "acute nephritis," and 15.8 per cent. as "simple peritonitis." There were 89 additions to "measles," 98 to "scarlet fever," 45 to "whooping cough," and 35 to "diphtheria and croup," making an addition of 267 deaths to these four diseases of childhood. "Acute nephritis" played an important part in the original assignment in all of these four diseases, but more especially in "scarlet fever," where 68 deaths, or 69.4 per cent. of the changes, were originally so classified. In like manner, "bronchopneumonia" and "pneumonia" (unspecified), were the original assignments in many of these conditions; this applies particularly to "measles" and "whooping cough." Except for "diphtheria," the indefinite title "simple meningitis" is a convenient cloak for the acute infections of childhood. "Influenza" and" tuberculosis of the lungs" both received significant additions from the title "pneumonia," the former receiving 228 additional deaths and the latter 70. Of the total changes to tuberculous meningitis, 294, or 78.0 per cent., were originally classified as meningitis.

An analysis of the sources of the additions to syphilis is interesting. Of the 1,031 changes to this title, 148 were from locomotor ataxia, 491 from general paralysis of the insane, and 104 from a large number of conditions, connected for the

Digitized by Google

most part with spinal lesions. Our inquiries with reference to "general paralysis of the insane" have definitely confirmed the virtually unanimous opinion of clinicians that this condition is syphilitic in origin. So uniform is our finding, that we are considering seriously the advisability of discontinuing our inquiries into this condition, and assuming the existence of syphilis in such cases.

As might be expected, the additions to "cancers of specified organs" are at the expense of the residual title, "cancers of organs not specified." The importance of this change is obvious in view of the increasing interest in the location of the malignant growths. The number of deaths from "acute and chronic alcoholism" was increased at the expense of "acute nephritis"; 213 out of 289 deaths, or 73.7 per cent., were so added. It will be unnecessary to go further into this analysis, since the various parts of the table speak for themselves.

The foregoing tables and text are sufficient to indicate the possibilities inherent in the intensive questioning of mortality returns. It does not lie within the power of the statistician to correct errors which depend upon the diagnostic ability of physicians, or upon circumstances which do not permit adequate opportunity for case observation or for accurate clinical or pathological findings. There is, however, a class of errors which he can undoubtedly reduce to a minimum. He can raise the standard of accuracy in mortality statistics by inviting the physician to give information such as he is able and willing to impart. It is reasonable to assume that the physician will, at first, have in mind not so much the statistical use of the information which he possesses as the immediate and uppermost clinical phase of the case. His efforts in the last stage of a losing battle against disease or injury are so often devoted to combating serious complications, that he is more than likely to certify only the immediate and not the primary cause of death. The statistician can tactfully point out to him an error of omission. Experience has demonstrated that, in a large proportion of cases, physicians manifest sufficient interest in the precision of our vital statistics to supply the additional and qualifying statements. As inquiries are made, moreover, physicians become more and more aware of the requirements

Digitized by Google

of statistical technique in making assignments. An analysis of our returns shows an increasing proportion of satisfactory statements for certain causes of death. A physician to whom an inquiry is once sent will usually not make the same indefinite statement in his subsequent certificates of death.

This means for advancing the accuracy of cause of death statistics is at hand in every registration and compiling office, and it is hoped that it will soon be adopted by vital statisticians generally. The fruitful character of the returns, affecting as they do the basic material of vital statistics, the simplicity of the method, and its low cost both in time and in money, indicate plainly the desirability at the present time of inquiries to physicians as an integral part of the classification of causes of death. Certainly the day has arrived when material cannot be admitted into the realm of good vital statistics unless it has previously been subjected to careful inquiry such as we have outlined above. That is the least which can be expected, in view of the uncertainty of much of the material, due in the first instance to errors of bedside diagnosis uncorrected by autopsy.

Digitized by Google

STATISTICAL TABULATION AND PRACTICE.*

By Roland P. Falkner, Ph.D., *Alexander Hamilton Institute.*

Nowadays the use of statistics is widespread and resort to tables or to graphic representation permeates modern literature on all subjects. There seems scarce a subject under heaven which someone does not present to us in tabular form or seek to make clear to us by means of charts and diagrams. What was at one time the peculiar reserve of the statistician and the economist has become the hunting ground of all the sciences, the pseudo sciences, and the advertiser. Statistics have become a common language. Into this language, whether in tables of figures or in graphic forms, everyone ventures boldly without even a Meisterschaft or a correspondence course preparation.

Most of us have blundered into it, and even those who may with best right claim to be experts, would admit that with them statistics is more of a habit than a principle. The result is an extreme diversity in habits of expression, which defeats the first purpose of statistical methods, whether in figures or diagrams, which is clearness of expression for groups of facts.

It is perhaps in the field of graphic presentation that the tyro indulges in the wildest flights of misrepresentation. The search for the striking and unusual leads to a perversion of fact. But when there is not conscious effort, inexperience often leads to confusing and misleading attempts to portray facts graphically. Earnest efforts are being made by the joint committee whose report was printed in the December issue of the Quarterly Publications to establish definite standards and, if they succeed in this laudable endeavor, much misdirected energy should be turned into profitable channels. Very little has been done in the search for standards in constructing and printing statistical tables.

The problems of tabular presentation of facts are comparatively simple and have not consciously been faced by

* Paper presented at the annual meeting of the American Statistical Association, Washington, D. C., December 29, 1915.

Digitized by Google

many statisticians. For that reason, they are also somewhat technical, but they are none the less real. The object of this paper is to present certain considerations in regard to arrangement and use of tables, with a view to raising the question whether standardization is necessary or advisable.

Problems of tabular presentation have had little consideration outside of government bureaus of statistics, each of which in the interest of uniformity in its publications, and wholly independent of the others, has developed rules governing the practice to be followed. There seems to be no reason why a like need should not be felt by the larger research foundations and by the statistical journals.

The ordinary printed table, perhaps not more than four or five columns wide and oftentimes not more than half as deep as the page on which it is printed, offers little difficulty. The eye takes in all its parts at a glance and the arrangement is not of vital importance. Difficulties begin only with larger tables. These we shall consider, and if rules for their construction were to be adopted, it would seem reasonable to apply such rules to smaller tables as well.

Without pretending to exhaust the infinite complexity of facts and their relations which may be expressed in tabular form, our attention will be directed to certain frequently recurring types which it is hoped will sufficiently emphasize what appear to be underlying principles of table construction.

The Whole and Its Parts, Stub Arrangement. One of the most frequent series of facts to be recorded is when facts are given for some total, say a geographic unit and its component parts, these being arranged in the stubs (Exhibit A).

The common arrangement is to name the parts first and place the total at the bottom (Table I). The reason given for such an arrangement is usually that in a sum in addition, the total is placed at the bottom. This would be conclusive if the addition were the main fact presented in the table, but it is obviously subsidiary. The chief interest attaches to the relations of the parts to the whole, and as there is one factor common to all these relations, the total, this is the most conspicuous fact to be exhibited.

Digitized by Google

EXHIBIT A.

POPULATION, AREA AND DENSITY, NEW ENGLAND, 1910.

Alternative Arrangement.

I.

Division and State	Population	Area in Sq. Miles	Population per Sq. Mile
Maine	742,371	29,895	24.8
New Hampshire	430,572	9,031	47.7
Vermont	355,956	9,124	39.0
Massachusetts	3,366,416	8,039	418.8
Rhode Island	542,610	1,067	508.5
Connecticut	1,114,756	4,820	231.3
New England	**6,552,681**	**61,976**	**105.7**

II.

Division and State	Population	Area in Sq. Miles	Population per Sq. Mile
New England	**6,552,681**	**61,976**	**105.7**
Maine	743,371	29,895	24.8
New Hampshire	430,572	9,031	47.7
Vermont	355,956	9,124	39.0
Massachusetts	3,366,416	8,039	418.8
Rhode Island	542,610	1,067	508.5
Connecticut	1,114,756	4,820	231.3

It is not infrequent that we desire to display the whole in relation to *some* of its component parts, *e. g.*, total wheat production in bushels and most important wheat producing states, without *all* parts. Here the analogy with a sum in addition fails entirely.

If in such a table, the center of interest is the total, in our illustration, New England, it cannot fail to be noted that this arrangement places the most conspicuous fact in the poorest position in the table as respects legibility. If a table extends over a page, this disadvantage is greatly emphasized. Many tables of the Statistical Abstract require the book to be turned up and fill two pages with the total at the bottom of the second page. Here a poor adjustment of the pages further complicates the situation. One table given here often covers 30 pages with the total on the last page. (Table 237, 1913, pp. 412 to 441. Exports of domestic merchandise.)

The census usage places the total at the top (Table II). This obviates all the disadvantages of the common arrangement. It places the most conspicuous fact in the posi-

Digitized by Google

tion where it is most easily read whether the table occupies two inches, a page, or several pages. It represents a finished product, not a process. Hence any conclusions drawn from the process of addition do not apply. The parts displayed need not under this arrangement equal the total. Parts may be omitted, or parts may be subdivided as for instance where in a list of states it may be desirable following New York State, to give the same facts for New York City.

The Whole and Its Parts, Box Arrangement. Convenience of arrangement not infrequently dictates placing the whole and its parts in the box heads rather than the stub (Exhibit B.)

EXHIBIT B.

PER CENT. OF TOTAL POPULATION IN CERTAIN AGE CLASSES, NEW ENGLAND, 1910.

Alternative Arrangements.

I.

Ages in Years	Me.	N. H.	Vt.	Mass.	R. I.	Conn.	New England
Under 5.....	9.7	9.2	9.6	9.8	10.0	10.1	9.8
5–14.....	17.7	17.0	18.0	17.2	17.6	17.7	17.4
15–24.....	17.1	17.4	16.8	18.5	19.5	18.8	18.3
25–44.....	28.2	29.2	28.5	32.5	32.0	31.7	31.4
45–64.....	18.9	19.2	18.8	16.7	16.2	16.3	17.1
67 and over.....	8.2	7.9	8.2	5.2	4.6	5.3	5.9

II.

Ages in Years	New England	Me.	N. H.	Vt.	Mass.	R. I.	Conn.
Under 5.....	9.8	9.7	9.2	9.6	9.8	10.0	10.1
5–14.....	17.4	17.7	17.0	18.0	17.2	17.6	17.7
15–24.....	18.3	17.1	17.4	16.8	18.5	19.5	18.8
25–44.....	31.4	28.2	29.2	28.5	32.5	32.0	31.7
45–64.....	17.1	18.9	19.2	18.8	16.7	16.2	16.3
65 and over.....	5.9	8.2	7.9	8.2	5.2	4.6	5.3

Here again alternative arrangements are possible. Frequently the whole is placed last in the column (Table I) and this follows similar impulses which led to putting the total at the bottom of the stub. We read from left to right and if the total is regarded as the result of the parts the last column is the logical place for it. Cross addition is less frequently used than up and down addition, but when it is used, it generally proceeds from left to right.

Digitized by Google

But if we emphasize the fact that a statistical table is a result and not a process, all the weight of argument, whatever may be the usage, favors placing the total in the column following the stub (Table II). Here it can be most easily read, and as it is often the central fact around which the table revolves, it is entitled to the best place.

Sequences, Stub Arrangement. A frequent type of statistical table shows similar facts for a sequence of years or months, the facts being recorded in the boxes, the years or months in the stubs (Exhibit C).

EXHIBIT C.

DENSITY OF POPULATION, NEW ENGLAND, 1860–1910.

Different Arrangements.

I.

Year.	New England.	Me.	N. H.	Vt.	Mass.	R. I.	Conn.
1860.....	50.6	21.0	36.1	34.5	153.1	163.7	95.5
1870.....	56.3	21.0	35.2	36.2	181.3	303.7	111.5
1880.....	64.7	21.7	38.4	36.4	221.8	259.2	129.2
1890.....	75.8	22.1	41.7	36.4	278.5	323.8	154.8
1900.....	90.2	23.2	45.6	34.7	349.0	461.6	188.5
1910.....	105.7	24.8	47.7	39.0	418.8	508.5	231.3

II.

Year.	New England.	Me.	N. H.	Vt.	Mass.	R. I.	Conn.
1910.....	105.7	24.8	47.7	39.0	418.8	508.5	231.3
1900.....	90.2	23.2	45.6	37.7	349.0	461.6	188.5
1890.....	75.8	22.1	41.7	36.4	278.5	323.8	154.8
1880.....	64.7	21.7	38.4	36.4	221.8	259.2	129.2
1870.....	56.3	21.0	35.2	36.2	181.3	203.7	111.5
1860.....	50.6	21.0	36.1	34.5	153.1	163.7	95.5

III.

Year.	New England.	Me.	N. H.	Vt.	Mass.	R. I.	Conn.
1910.....	105.7	24.8	47.7	39.0	418.8	508.5	231.3
1860.....	50.6	21.0	36.1	34.5	153.1	163.7	95.5
1870.....	56.3	21.0	35.2	36.2	181.3	203.7	111.5
1880.....	64.7	21.7	38.4	36.4	221.8	259.2	129.2
1890.....	75.8	22.1	41.7	36.4	278.5	323.8	154.8
1900.....	90.2	23.2	45.6	37.7	349.0	401.6	188.5
1910.....	105.7	24.8	47.7	39.0	418.8	508.5	231.3

The usual order places the earliest date at the top and the latest at the bottom. As we read a page from top to bottom,

Digitized by Google

we naturally expect a sequence in this order (Table I). While this arrangement corresponds with the normal order of a sequence it has the disadvantage of placing the last date in the most inconspicuous position.

A contrary arrangement familiar in census publications places the latest date at the top and the earliest date at the bottom (Table II). Perhaps all of us have seen men who were not figuratively, but literally as bald as a billard ball. It disconcerted us at first, but in time we got used to it. This table produces the same effect. Constant use may familiarize us, but we never quite forget the shock of first acquaintance. Its awkwardness remains and we constantly find ourselves reading down instead of up, and a conscious effort is necessary to get at the true meaning.

None the less, the census usage is not a monstrosity but well founded. It should be remembered that the primary interest in census figures is the picture which they give of present conditions and these are emphasized when the latest figures are given the first place.

The reason for such totally different practices lies in the distribution of emphasis. What is most important, the sequence or the present status? The first arrangement emphasizes the sequence, the second the present status. Both of these factors of importance enter into any figures of this character.

In this apparently insoluble conflict, it may be that the most satisfactory usage would be a compromise which repeated the latest year and gave it both the first and the last place (Table III).

Sequences, Box Arrangement. When the years or months appear in the boxes, a somewhat similar situation arises (Exhibit D).

If the normal sequence from left to right is followed (Table I), the last date comes in the last column in the least conspicuous part of the table where it is least easily read. In so far as the interest centres in present conditions as in census enumerations, this is a distinct disadvantage.

Digitized by Google

EXHIBIT D.

DENSITY OF POPULATION—NEW ENGLAND, 1860–1910.

Different Arrangements.

I.

Division and State.	1860.	1870.	1880.	1890.	1900.	1910.
New England	50.6	56.3	64.7	75.8	90.3	105.7
Maine	21.0	21.0	21.7	22.1	23.2	24.8
New Hampshire	36.1	35.2	38.4	41.7	45.6	47.7
Vermont	34.5	36.2	36.4	36.4	37.7	39.0
Massachsetts	153.1	181.3	221.8	278.5	349.0	418.8
Rhode Island	163.7	203.7	259.2	323.8	401.6	508.6
Connecticut	95.5	111.5	129.2	154.8	188.5	231.3

II.

Division and State.	1910.	1900.	1890.	1880.	1870.	1860.
New England	105.7	90.3	75.8	64.7	56.3	50.6
Maine	24.8	23.2	22.1	21.7	21.0	21.0
New Hampshire	47.7	45.6	41.7	38.4	35.2	36.1
Vermont	39.0	37.7	36.4	36.4	36.2	34.5
Massachusetts	418.8	349.0	278.5	221.8	181.3	153.1
Rhode Island	508.5	401.6	323.8	259.2	203.7	163.7
Connecticut	231.3	188.5	154.8	129.2	111.5	95.5

III.

Division and State.	1910.	1860.	1870.	1880.	1890.	1900.	1910.
New England	105.7	50.6	56.3	64.7	75.8	90.3	105.7
Maine	24.8	21.0	21.0	21.7	22.1	23.2	24.8
New Hampshire	47.7	36.1	35.2	38.4	41.7	45.6	47.7
Vermont	39.0	34.5	36.2	36.4	36.4	37.7	39.0
Massachusetts	418.8	153.1	181.3	221.8	278.5	349.0	418.8
Rhode Island	508.5	163.7	203.7	259.2	323.8	401.6	508.5
Connecticut	231.3	95.5	111.5	129.2	154.8	188.5	231.3

On the other hand, the arrangement of dates so that the latest immediately follows the stub and the earliest comes in the last column (Table II) as in the census publications, offends our natural instinct to read from left to right and interferes with the ease with which the sequence is comprehended.

Here again a compromise between conflicting interests may be suggested by giving first the latest date, giving earlier ones in the natural order, and repeating for the sake of the sequence the last date in the last column (Table III).

Digitized by Google

Mechanical Aids to Clearness. The object of the statistical table is to compress into a small space a body of information the narration of which in the text would be cumbersome and exhausting to the reader. It is in short a method of condensation, and it is of the utmost importance that, as it tells so much in so small a compass, it tell it as clearly as practicable. Statisticians cannot and others should not, therefore, neglect the various mechanical aids to clearness which are available. These are chiefly differences in type and rules. The total should always be distinguished from the parts either by the use of heavier type, or by a rule or space separating it from the parts, or by both methods. When the total is in the box head, a double rule or heavy rule is advisable.

In a chronological series emphasis can be placed upon the latest date by similar methods and this is often times desirable.

Numbers and Titles for Tables. The usual practice is to give each table a number and no other method affords such a simple plan for identification in the text comment. It has sometimes been argued that numbering tables makes them a thing apart from the text instead of part and parcel of the treatment of subject. Without contesting the plausibility of this argument, it must be admitted that a table of figures is just as much and no more a part of the text than is a map in any geographical publication. It is somewhat less distinct perhaps than an illustration, but a certain separateness cannot be denied and cannot be avoided by ignoring it. The omission of table numbers leads to difficulties in practice which it is the first duty of the statistician to avoid.

Whether tables should have titles as well as numbers is largely a question of taste. Larger tables commonly have such titles, but they are often omitted in text tables. If the title is helpful in text references, it should not be omitted. My personal preference is for the use of titles in every instance no matter how large or how small the table. In any case, long titles are to be avoided. The practice of some government bureaus of giving in the title all the elements which enter into the box heads, involves useless repetition. When, for any reason, long descriptive titles are felt to be necessary, a compromise may be effected by the use of a short catch title, followed by the descriptive title.

Digitized by Google

Contents of Tables, Abbreviated Statements. The contents of statistical tables vary so much according to the nature of the problems involved, that it might seem that we should look in vain for any general rules. However, one type of problem constantly recurs wherever round numbers or percentages are used. Statistical accuracy is often served by giving such larger items as wheat production, value of farms, manufactured products in terms of millions of dollars. Yet our practice in such cases is to grasp after even greater exactness. We may give one decimal after the millions as notice perhaps that the numbers are not round after all. Then we add up the abbreviated figures and carefully adjust the decimals to see that they add up exactly to the sum given in the detailed figures.

Again we frequently have a division of some total into percentages. These again are often given with one or two decimals attached. Again we are obsessed with the idea that the sum of such percentages should equal 100 and we painfully "adjust" them to bring about such a result.

In any proceeding of this kind, we are obliged in some measure to twist facts to make them suit the table. If we were to really grasp the idea that in presenting a statistical table, we are giving not a sum in addition but a finished result, we should not expect our tables to prove. Whether the sum of our percentages equalled 99.7 or 100.5, each item would be correctly stated and we should have done no violation to our conscience in doctoring the figures to get the desired result.

Conclusion. A few of the questions which arise in the practice of statistics have been discussed in this paper. To those who are not at least in some degree versed in the subject, they might appear trivial and inconsequential. But to those who are interested in perfecting the precision of their instruments of thought and expression, they have their significance. If the aim of all statistical statements is clearness and sanity we cannot neglect anything which contributes to this end.

Digitized by Google

STATISTICAL STANDARDIZATION IN WASHINGTON.*

By Julius H. Parmelee, Ph.D., *Statistician, Bureau of Railway Economics.*

Dr. Falkner has wisely handled the important subject of standardization in a concrete and helpfully suggestive fashion. I am glad of the opportunity of continuing the good work, if possible, by adding a few concrete suggestions regarding the problem, pointing my remarks with especial reference to conditions in Washington.

The field of government work, here in Washington, offers perhaps the best opportunity for standardization of statistical effort. Here are a number of federal bureaus, of which the Census Bureau is the best example, devoted exclusively or principally to the compilation of statistics, while thirty additional bureaus and offices are engaged more or less extensively in statistical work. Some of these bureaus cover fields that overlap, but although there exists at all times the danger of statistical duplications, or even contradictions, I think it safe to say that the danger has been a continually diminishing one during the past few years.

Turning specifically to the problem of standardization as applied to the statistical work of the government, we find that in September, 1908, the President appointed an Interdepartmental Statistical Committee of ten members under the jurisdiction of the Secretary of Commerce and Labor. This committee was selected mainly from among the government statisticians of Washington. Among other things, the committee was directed to make recommendations:

(1) With a view of establishing uniform definitions of statistical terms.

(2) With a view of introducing uniform methods.

(3) In regard to schedules and accompanying instructions relative to any new line of statistical inquiry; above all, to test the results achieved, and to investigate whether they are

*Paper presented at the annual meeting of the American Statistical Association, Washington, D. C., December 29, 1915.

Digitized by Google

in harmony with each other and modern statistical methods and practice.

These directions, it will be noted, go to the very heart of the problem of standardization. When uniformity has been achieved in definition of terms and in methods, and harmony with the best statistical practice has been attained, the goal of standardized presentation is not far removed.

However, this goal too nearly approached the ideal to be gained at once. The committee, together with a subcommittee on duplication of government statistics, held a number of sessions, and issued two or three reports. The investigations of the two committees traversed the whole field of statistical work in Washington, but a large part of the discussion centered around the advisability of establishing a central statistical office to serve as a clearing house for all the departments and bureaus of the federal government.

So far as I know, the matter went no further than these reports. Technically, however, the Interdepartmental Committee is still in existence, and could once more be ordered to conduct investigations in the direction of standardization, if the proper initiative were taken. In other words, the machinery for setting standardization methods in motion is still existent.

Before turning to a concrete instance of statistical standardization and its needs in the federal organization, let me point out that a government bureau should first of all standardize its own work. It ought not to be possible to say that an important bureau in Washington tabulates statistics of retail prices and of wages and hours of labor, summarized on such different forms as to make it almost futile to bring them together, except in a most limited way; yet this is the fact. Again, the different bureaus should coöperate with each other to attain the greatest degree of uniformity and standardization. This is being done by means of informal conferences and agreements, the appointment of such committees as the Interdepartmental, and the like. In the third place, there should be greater coöperation, necessarily involving a certain amount of standardization, between the federal government on the one hand and state and local governments on the other hand.

Digitized by Google

Perhaps the best illustration of the crying need for statistical standardization of this latter sort grows out of the requirement of periodic or occasional reports of the operations of public service corporations, especially railways. Development of this method of securing statistical reformation for purposes of regulation has been extremely rapid. A certain Eastern railway system, operating through eight states, is required to make over 144,000 separate reports each year to federal, state, and local officials. These reports cover every conceivable phase of finance and operation. Many of the reports cover the same, or similar, ground; yet the blanks or forms provided for the reports are not always uniform as between federal and local requirements. The Interstate Commerce Commission has been doing good work in the direction of standardization by furnishing public service commissions of various states with blank forms similar to those on which the Interstate Commission itself requires reports from steam and electric railways, from express companies, and from sleeping car corporations. These forms are utilized by many state commissions for the reports which they require from railroads or other public service corporations operating in their respective states. More than four fifths of the states now use the Interstate Commerce Commission standard form for steam railways. The saving in energy and expense to roads operating in several states is clearly considerable. In addition the Interstate Commission furnishes standard forms to three fifths of the states for reports from electric railways, to two thirds of the states for express company reports, and to ten states for sleeping car company reports.

Yet even here there is further need for standardization. While forty state commissions require annual reports from the railways on the standard Interstate Commerce Commission forms, the statistical tabulations prepared by them on the basis of those same standard forms are anything but standard. In fact, even a hasty glance through the latest statistical reports of the state railway commissions—and regulatory commissions exist in some form in 46 of the 48 states of the Union —discloses almost as many varieties of statistical presentation as there are reports.

Digitized by Google

I do not wish to be understood as arguing that absolute uniformity of statistical presentation is desirable in these state reports. It is always a question, in fact, how far standardization may be attainable without the sacrifice of special interests and within the requirements of individual states. There exists room for desirable uniformity, however, in certain fundamental features of statistics as presented. Certainly, any one who has ever attempted to utilize the statistical tables of two or more state railway commissions knows how impossible it is to bring them into line with each other.

Clearly no better opportunity exists for improvement of statistical work, especially in the direction of uniformity and standardization, than in this field of railway statistics. The reports are rendered on a standard form, the first step in the process of standardization, but their compilation is on anything but a uniform or standard basis. I would suggest to members of the two Associations meeting here this morning that there is no better field for their efforts than to labor for greater uniformity, and consequent serviceability, in the presentation of statistics of the transportation industry.

Digitized by Google

CLASSIFICATION AND LIMITATIONS OF STATISTICAL GRAPHICS.

By E. A. Goldenweiser.

Graphic methods of presenting statistical facts have been in use from times immemorial and are doubtless destined to become increasingly popular. Mr. Willard C. Brinton's book on Graphic Methods of Presenting Facts (*The Engineering Magazine*, 1914) and the subsequent creation of a committee of representative men from various fields of endeavor for the study and standardization of graphics are indications of the increasing interest in the matter and are themselves in part responsible for this increased interest. The committee has made a preliminary report* in which it lays down seventeen elementary rules to be followed in presenting facts graphically. Considering the wide divergence of interests on the part of the members of the committee, which includes engineers, economists, statisticians, psychologists, mathematicians, accountants, educators, and others, the report must be considered as a substantial achievement and a decided step in the direction of standardization. The great need of standardization was perhaps the principal fact impressed upon the visitor to the exhibit of graphics held under the auspices of the committee in connection with the sessions of the Second Pan-American Congress and of the various scientific societies that met in Washington, the last week of December, 1915.

It is the object of this paper to contribute to the work of standardization by attempting to classify the uses of graphic methods and to delimit briefly their fields of greatest usefulness. The discussion is confined to statistical graphics and does not touch upon devices like organization charts, routing charts, and the like, the subject matter of which is not statistical.

Statistical graphs may be classified in the first place according to method of presentation and secondly in relation to primary object. The classification on both of these bases might be presented as follows:

* See Quarterly Publications of the American Statistical Association, December, 1915, p. 790.

Digitized by Google

STATISTICAL GRAPHICS.

Method of presentation.	Object.
I. Publication in books or bulletins.....	1. Reference
	2. Illustration
	3. Analysis
	4. Research
II. Use at lectures.....	1. Reference
	2. Illustration
	3. Analysis
III. Use in exhibits.....	1. Reference
	2. Illustration
	3. Analysis

The discussion will take up the graphs devoted to each of the four primary objects and will refer to the methods of presentation only incidentally. It may be noted at the start that graphs are often, if not generally, used for more than one of the objects mentioned, and quite properly so; nevertheless, a clear understanding of the primary object for which the graph is prepared will be helpful in deciding on its character and arrangement.

1. Reference Graphs. A graph, like a general table, may be prepared with no other object than to present a given set of facts. A curve showing the value of exports during a series of years is such a graph; a series of forty-eight bars showing the population of each state is another; and a map showing, by means of dots, the distribution of a crop is a third. The map, however, contains certain elements of correlation: the geographic location and concentration of the crop stands out; it appears whether the crop is in the South, the North, or the West; furthermore, a common knowledge of the location of the principal cities and rivers superimposes a mental set of correlations, even if the map is actually shown in bare outline. If the map shows isotherms, isobars, or topographic contours, it becomes an analytical graph. As a general proposition reference graphs are seldom of value unless they carry a time series, like the curve, or a geographic relationship, like the map. In other words, their value is dependent on their being more than pure repositories of facts. The bar

Digitized by Google

diagram mentioned would probably be valueless unless the states were arranged in descending order, thus introducing a relationship. It may be said, therefore, that reference graphs are to be used with caution; they are often a waste of time and space. If used at all, reference graphs must be absolutely simple and clear, for their only excuse is that they may be more easily comprehended by eye-minded readers than a statistical table which is generally more compact, more accurate, and more easily handled.

Graphic methods are most effectively applied not to raw material but to results. There would be little excuse for stating this were it not for the fact that numerous graphs are published that have no apparent object other than to reproduce unassorted facts. Such graphs may be of some use in a lecture by helping the lecturer to pin his words to a concrete exhibit, but in a printed publication or an exhibit a graph should never be shown unless it helps bring out some significant relationship more clearly than do the text and tables.

2. Illustrative Graphs. An illustrative graph differs from a reference graph in that the former is selective while the latter is inclusive. A bar diagram comparing the corn production of Illinois with that of the United States is illustrative; so is a series of bars showing the number of deaths from given causes. An illustrative graph is very useful in print, as well as in lectures or exhibits, because it helps to fix in the mind of the reader some important fact that may be the key-note of a discussion. Such a graph should be very simple since it is an appeal to visual memory, and if, in order to remember correctly, one needs to keep in mind many items and qualifications that are not impressed on the eye, the graph loses a considerable part of its usefulness. An illustrative graph may be comparatively complex and not self-explanatory if it is intended for use in connection with a lecture where the speaker has the opportunity of pointing out the significance of the chart. Such a graph serves a somewhat different purpose from that of graphs on a printed page: it has a psychological value in giving the audience something to look at and helps hold their attention. This advantage

Digitized by Google

is forfeited, however, when the lecturer has too many charts, or exhibits more than one at a time, or spends too much time in adjusting one after the other. In such cases the attention of the audience is diverted from rather than attracted to the discussion.

3. Analytical Graphs. An analytical graph is seldom used except in connection with printed text or a lecture; at an exhibit it would demand too much explanation. An analytical graph is one that is introduced into a discussion in order to show visually a relationship that the author wishes to emphasize. A graph consisting of two curves—one showing the value of the potato crop through a series of years and the other the production of the same crop—is analytical. The two curves are brought together in order to emphasize the circumstance that a large crop often results in so low a price that the total value of the large crop is smaller than that of a smaller crop harvested during another year. Such a graph is of distinct value because the correlation is made much plainer by the curves than it can be made by text and tables alone. A map showing mean summer temperature and the distribution of a crop is analytical in that it exhibits the temperature control of the crop. Another illustration of an analytical graph is shown in the accompanying diagram, taken from a bulletin of the Department of Agriculture.*

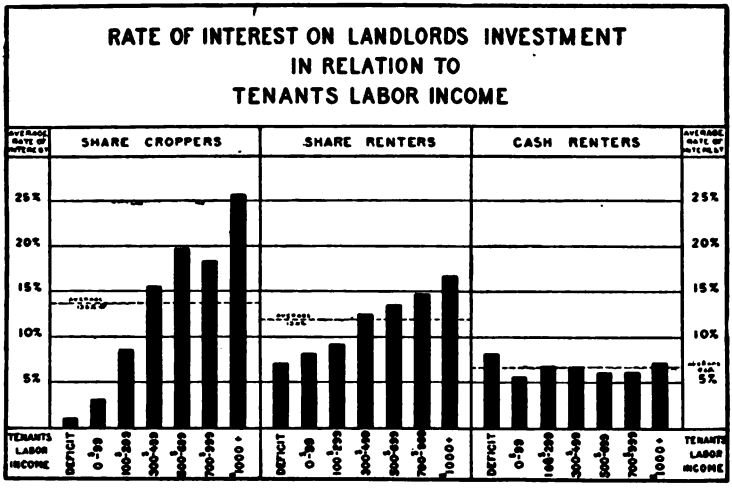
The fact that the landlord's rate of interest varies greatly in one class of tenure in conjunction with the tenant's income, while in the second class the variation is much less, and in the last class it is insignificant, stands out more clearly from this graph than from any amount of discussion. One may perhaps be justified in the belief that it is in the field of analysis that lies the greatest usefulness of statistical graphics. A graph of this sort not only visualizes a conclusion previously reached but even helps to reach the conclusion and thereby approaches the last type of graph to which we now turn.

4. Research Graphs. A research graph is one that helps establish an unknown correlation as a result of plotting ascertainable data. This type of graph, used largely in the laboratory and in print, is more common in engineering and natural

* U. S. Dep't of Agr. Bulletin No. 337, p. 13.

Digitized by Google

science than in statistics, but statisticians sometimes receive hints of possible relationships from experimental research graphs. These graphs are not, like the others, primarily a method of presenting, but of discovering a relationship. This highest form of graphics lies on the fringe of the subject of this paper and may be respectfully dismissed with these few words.



In conclusion, it may be stated that the great need in statistical graphics, as in statistics, is not for greater output but for higher quality. Standardization of method is highly desirable, but it is essential for statisticians to understand clearly that graphic methods for them are merely subsidiary and are valuable as a rule only in so far as they help to popularize the results of study. A transformation of a table into a graph, and especially of a good table into a bad graph, is not a statistical achievement. The statistician should not permit his interest in the graphic presentation of his material to divert him from his main work of careful statistical interpretation, the insufficiency of which is today the crying need in the study of social conditions.

Digitized by Google

THE USE OF THE CORRESPONDENCE METHOD IN ORIGINAL RESEARCH.

By Asher Hobson, *University of Wisconsin.*

This paper is neither a defense nor an advocacy of the correspondence method of conducting original research. It is an analysis of the process with a view to understanding its best use in those cases where the personal-interview method cannot be employed. This analysis is based upon the results of the experience of the Department of Agricultural Economics at the University of Wisconsin, in the use of correspondence in conducting original research studies in the field of marketing. The department has tried different types and forms of letters in gathering the same sort of data, and has tabulated the results from the standpoint of efficiency—number and reliability of answers. While the conclusions drawn may not hold true in all lines of original research, they must be applicable, generally speaking, to the field of economics.

As compared with the personal interview method of gathering data from original sources, correspondence is open to much criticism. On the whole it can safely be said that the correspondence method is unsatisfactory because of the unreliability of the data obtained in this manner. The indefiniteness and inaccuracy of answers, which in most cases are due to misinterpretation of questions, give rise to a large percentage of error. But with all its faults, its use is indispensable to the investigator in gathering certain classes of data. First hand information which pertains to a wide area or data concerning segregated districts separated by great distances can only be obtained, without the expenditure of large sums, by correspondence. There are times, therefore, when correspondence is the only agency to be considered, and the vital question is how to eliminate the errors resulting from its use.

Nature of Questions. Statisticians state the truism that questions should be few in number, very simple and easy to answer. But the problem is not solved by that formula; for

Digitized by Google

the difficulty arises on the one hand in attempting to make the questions few enough, simple enough and easy enough to answer in order that the replies may be definite and trustworthy, and on the other in embodying enough in them to cover the field desired.

One sometimes wonders whether the required simplicity can be attained in all cases. It is surprising from how many different angles a seemingly simple question will be answered. An inquiry addressed to wholesalers over the country as to the wholesale price of cheese for a certain day or week would quite likely result in data pertaining to (1) the wholesaler's selling price, some quotations including freight and some not, (2) the wholesaler's buying price with freight charges considered in some cases, (3) and moreover different answers would pertain to different kinds of cheese, while (4) some replies would refer to foreign and some to domestic varieties of the same kind. It is safe to predict that the one question would be answered in many different ways.

The results of the experience of this department in its attempt to overcome the above difficulty point to the superiority of what might be called a one-question letter. This type of letter certainly has tendencies toward simplicity in that the questions are reduced to a very few—quite often one—which allows for minute qualifications.

Another advantage in asking only one or a few questions is that it tends to overcome that antagonistic spirit encountered by so many economic investigations. Prejudice is a barrier which the economist must always be prepared to meet. Many of his problems pertain to personal matters, private business and stocks in trade, so to speak. Because of these prying propensities he is looked upon with suspicion. To ask a business man for his handling margins is little short of impertinence, to say nothing of the probability of his answer being a prejudiced one. But to ask one merchant for his buying price under certain conditions, and another merchant for his selling price under the same conditions is a different matter. Questions which are abstract and impersonal and not connected with a series of other questions will be rewarded by more and franker answers. Hence the one-question letter,

Digitized by Google

like the example below, goes a long way in overcoming that everpresent none-of-your-business attitude.

MADISON, WIS., May 2, 1914.

GENERAL GROCER CO.,
CHICAGO, ILLINOIS.

Gentlemen:

The Wisconsin Agricultural Experiment Station is making an effort to collect all the information possible on the subject of marketing. At present we are studying the marketing of cheese. If you will kindly fill in the blank below and return it to us, your coöperation will be greatly appreciated.

Yours very truly,
WISCONSIN AGRICULTURAL EXPERIMENT STATION,

Per_____

Wholesale Selling Price of Cheese, f. o. b. your station—for *last week* in March, 1914.

In cents per pound.

American.	Swiss.	Brick.	Limburger.
	Imported Domestic		

From whom do you buy your cheese?

Name of firm_____

Address_____

Use other side for remarks.

However, this type of letter has its drawbacks. If only one question is asked when the answers to five are wanted, more letters will be required than if each letter had contained all five questions.

The chief claim for the correspondence method is its cheapness, yet no method is cheap that does not give fairly correct results. Hence the questionnaire manner of securing data is not cheap unless it can be made reliable. Postage alone costs $40 per 1,000 letters, including stamped return envelopes. Consequently, it is a loss of time and money to send out bulky, complex questionnaires which are likely to be incorrectly answered, or more often not answered at all. The greater

Digitized by Google

cost of the meager letter is offset in a large degree by a higher percentage of answers, to say nothing of the trustworthiness of the answers. The only absolute limitation on this type of questionnaire is a mailing list so meager as not to permit its division into different groups.

Verifying the Data. Any information, however secured, should be verified if possible. One means of doing this is by getting the same sort of data from different sources. For instance, take the question of middleman charges in a simple form. The producer's selling price is the same as the manufacturer's or dealer's buying price, the manufacturer's selling price is the jobber's buying price, and the jobber's selling price is the retailer's buying price. Each step affords two widely different sources for the same information. But the checking up of the retailer's selling price cannot be done in the same manner, since the consumer buys under such varying conditions that their buying price would be of little value as an index.

In such a case one must depend upon his own knowledge of the subject as to whether the answers are representative. Question the incentive toward misrepresentation. What motive would the correspondent have in misstating the facts? Answer this question by framing a letter that will reduce such a motive to a minimum. The principal idea of a few disconnected questions is to eliminate this wish to misrepresent. To ask a retailer for his selling price of a specific commodity is more apt to result in a reliable answer than if the request were accompanied by others pertaining to his business.

Data that cannot be verified by collection from different sources should be submitted to a critic who is in a position to be able to detect errors. This critic will be, generally, a man actively engaged in the field of business which is being investigated. But, on the whole, instances are rather rare in which given data cannot be verified by comparison with those from different sources.

Help the Correspondent. Avoid bulk. Letters should not only contain a very few questions, but also a very few pages—one if possible. A bulky questionnaire often looks more formidable than it really is and for that reason is thrown in the waste basket without serious consideration. A one-page

Digitized by Google

letter which can be answered on the same sheet on which it is written has a distinct advantage. (See sample letter p. 212.) Its appearance of simplicity, and ease in answering, invites immediate attention. There are no extra pages to mislay or lose. Moreover it saves asking for the name and location of the correspondent, which are quite essential in most investigations—the location especially—and which are often omitted by the correspondent even when requested. This omission is due in some cases to carelessness and in other cases to a dislike—conscious or unconscious—in having a direct connection between the information and its source. The salutation on the one-page letter which is returned with the answer, saves the investigator the trouble of asking for name and address and the correspondent the trouble of writing them. Even though they were cheerfully given, the investigator should ask no questions which he can answer himself.

But even one question may at times necessitate the use of more than one page. In such cases form blanks should be constructed in sufficient detail to allow filling in with the least possible exertion. Avoid the asking of general questions since they often involve lengthy explanations. Blanks should be so constructed as to make explanations as few as may be. It is sometimes possible for the investigator to analyze the question minutely, and so construct the blank that the answers can be given by a series of check marks.

The idea is not to require the correspondent to make one stroke of the pen that can be made for him.

Incentive to Answer. Answers to letters asking for information depend in most cases upon the good will of the one upon whom the request is made and who is not interested in the promotion of the investigation, hence feels himself under little or no obligation to answer. For this reason it is well for the investigator, if possible, to so word his questions as to shift the answering from a basis of mere accommodation to one of interest. An experience will serve as an illustration. The Department of Agricultural Economics at the University of Wisconsin made an attempt to secure information concerning the organization of Wisconsin cheese factories—whether coöperatively or privately managed—in order to construct a

Digitized by Google

map showing the location of each type. After two rather unsuccessful letters, the third was sent out to the factories to the effect that a map of the state was being constructed, showing the location of its cheese factories, and if they cared to be represented upon this map, they should "fill in the blank below immediately." The returns were gratifying.

Providing the results of the investigation are intended for free distribution, it is a good plan to offer to send a copy upon its publication to those who will indicate their desire for it in their answer. This necessitates a reply in order to put in an application for the results of the writer's work. The following is an illustration of such an incentive: "If you are interested in our work on cheese marketing we will gladly send you the bulletin immediately upon publication."

It is sometimes well to offer special inducements for replies. The following paragraph contained in a letter to Wisconsin creamerymen illustrates this point:

"In order to show our appreciation of an immediate reply we will mail you the average monthly price paid for butterfat by coöperative creameries in the state, and also the price paid by private companies, in order that you may compare them with your prices. Kindly indicate on the back of this page if you wish these figures."

The Make-up of a Circular Letter. A circular letter cannot easily be disguised, nevertheless its appearance should be that of a personal one. The expense of getting out circular letters by the use of the typewriter is prohibitive. The mimeograph process is cheaper, but its imitation of typewriting is so defective as to exclude its use except in writing to a class of people who do not readily differentiate between grades of correspondence. On the whole, the cheapness of mimeographing is offset by a decrease in replies. The better method is that of multigraphing. It costs slightly more than mimeograph work but is an exact representation of typewriting which permits the use of that machine in filling in specific salutations at the head of each letter.

Avoid the use of "Dear Sir"; or "Madam"; which may apply with equal appropriateness to the president of our country or the village blacksmith. Effective circular letters

Digitized by Google

must have an air of the personal. This requires a signature and not the use of a stamp.

To Whom to Write. The tendency is to write to those directly interested; that is, get farming information from farmers, and retailing data from merchants. But such procedure is not always best. In some cases those directly interested are the ones who should not be approached. Many people dislike to give information about their own business, but are quite willing to divulge the nature of some one else's business. A retailer will often tell of a wholesaler's methods, as he has experienced them, when he will not converse about his own business procedure.

In considering whom to approach, first enumerate the different classes who can furnish the information. Then write to those who will be least prejudiced against answering, and those who are best prepared to handle correspondence. This department has discontinued the sending of questionnaires to farmers if the information can be secured from other sources, because of the low percentage of answers. Their failure to reply is not due, in most cases, to an unwillingness to answer, but to neglect. The handling of correspondence is not included in their daily routine. Questionnaires are forgotten, misplaced, or lost. It is not an uncommon occurrence to receive answers from farmers a year after the request has been sent. Hence our policy is to write professional men concerning general agricultural matters of their communities. The banker is acquainted with practices and business methods while the postmaster also is pretty well versed on local affairs.

The Mailing List. It is a broad statement to say that no matter what the class of people with whom one wishes to correspond, there is a published mailing list of some sort or other which contains their names and addresses. It may be the city directory, the commercial club roll, the county tax list, a financial register or one of the numerous commercial gazetteers. It may be issued for individual cities, by counties, by states or for the whole United States. At least a published list of some sort is obtainable.

The problem is not in being able to secure the names of a certain class, but in choosing the ones in the class who can

Digitized by Google

furnish the desired information. For instance, a very complete wholesale grocery directory would be of little value in the investigation of wholesale prices of Swiss cheese, since only a small percentage of grocery jobbers handle this variety of cheese. A directory containing the names of all the farmers in a state would be of little use to an investigator working on dairy butter problems in that state, if only one out of every twenty farmers made dairy butter.

The most complete commercial directory is apt to be incomplete as regards the needs of a particular investigator. A state directory of retailers omits the merchant in the cross-roads town. A state directory of wholesalers designates those handling cheese, but it doesn't differentiate between the kinds of cheese. So another broad generalization can be made to the effect that no matter how complete a commercial directory may be the probabilities are that it will not suit all needs.

This brings one to the problem of making up a mailing list to suit individual requirements. An illustration will best explain the situation. Suppose an investigator who is making a study of the marketing of dairy butter. His commercial list does not contain the names of merchants located in small towns and they are the very ones who handle dairy butter. If he followed his farmer's directory he would have to write twenty letters in order to reach one farmer who made butter. As a result both lists are discarded. Since postmasters are not permitted to give out the names of their patrons, he writes to the station agents of rural towns briefly explaining his difficulty. Then when writing to the merchants whose names the agent submits, he in turn asks for the name of farmers from whom they buy their dairy butter.

Take a different situation—that in which a majority of the class do not handle the commodity in question—Swiss cheese and the wholesale grocer. In such instances it is sometimes necessary to obtain the names of the customers of those handling the commodity in its preceding stage. In order to secure a list of wholesalers handling Swiss cheese it would be necessary to obtain the names of those to whom the dealers sold. (Dealer is a technical term applied to a cheese middleman who intervenes between factory and jobber.) However, it is

Digitized by Google

a very difficult task to obtain such a list from a commercial firm. Information of that sort in the hands of competitors would be injurious, to say the least. For this reason one should not resort to this means of compiling a mailing list, unless it cannot be made up from other sources.

Knowledge of the Subject. By all means one should possess a fair knowledge of the field which his investigation covers before making out the questionnaires. This working knowledge is best gained by some field work, supported by supplementary reading. If the investigator has not the opportunity to acquaint himself with the field except by correspondence, then his proposed questions should be submitted for criticism to someone who does know the subject in question, in order that none but intelligent inquiries be included. To expose one's ignorance of a subject by asking insignificant or irrelevant questions that are matters of common knowledge to members of the trade, is to create a prejudice against the letter which forestalls reply.

Digitized by Google

REVIEWS AND NOTES.

CASUALTY ACTUARIAL AND STATISTICAL SOCIETY OF AMERICA.

On November 7, 1914, the first organization meeting of the Casualty Actuarial and Statistical Society of America was held at the City Club in New York City. Arrangements were made for founding a Society to study the statistical and actuarial science of casualty and social insurance. The following officers were elected:

OFFICERS AND COUNCIL.

President ..... I. M. RUBINOW*

Vice-Presidents ..... BENEDICT D. FLYNN*, ALBERT H. MOWBRAY*

Secretary-Treasurer ..... CLAUDE E. SCATTERGOOD*

Librarian-Editor ..... WINFIELD W. GREENE*

Other Members of Council

	Term Expires.
JAMES D. CRAIG	1915
R. K. ORR	1915
FREDERICK L. HOFFMAN	1916
ALBERT W. WHITNEY	1916

At the first dinner of the Society, informal addresses were made on the various aspects of casualty and social insurance. The first number of the proceedings of the Society contained the articles entitled "Scientific Method of Computing Compensation Rates," by I. M. Rubinow, and "Valuation of the Death Benefits Provided by the New York Compensation Law," by W. W. Greene and a collection of reviews and notes which included an extensive bibliography of current literature on the foundations of casualty and social insurance.

The second volume of the proceedings was devoted to a discussion of the immediate problems facing the Society and gave also a competent review of the subject of workmen's compensation reserves.

The third number was devoted to a presidential address by Dr. Rubinow on schedule rating and to a number of papers on the same subject by fellows of the Society.

The fourth number of the proceedings contained an index of the first volume and articles of general statistical interest as follows:

The Classification of Industries for Workmen's Compensation Insurance, by E. H. Downey.

Note on the Application of Recent Mathematical-Statistical Methods to Coal Mine Accidents, by Arne Fisher.

A New Graphic Method of Using the Normal Probability Curve, by Buckner Speed.

*Members of council, *ex-officiis*.

Digitized by Google

The fifth meeting of the Society was held in New York City on February 25 and 26, 1916, and the following papers were read and discussed:

Mortality from External Causes among Industrial Policyholders of the Metropolitan Life Insurance Company, 1911–1914, Louis I. Dublin.

Work of the Bureau of Accident and Health Statistics, Benedict D. Flynn.

Cost Accounting in Casualty Insurance, Claude E. Scattergood.

Compensation Cost of Industrial Diseases, James D. Maddrill.

Statistics necessary for Computing Compensation Rates, Edward Olifiers.

American Methods of Compensating Permanent Partial Disabilities, I. M. Rubinow.

Since the initial meeting of the Society the following changes have occurred in the membership of the Council: Richard Fondiller was elected Editor-Librarian and Messrs. Harwood E. Ryan and Joseph H. Woodward succeeded the retiring members of the Council whose terms expired in 1915.

Examinations will be held for admission to associateship and fellowship in the Society during the first week in May, 1916.

The Society's committee on workmen's compensation statistics has published a classification of industries for general use in the compilation of industrial accident and compensation cost statistics. The classification appears in the November, 1915, number of the Monthly Review of the United States Bureau of Labor Statistics. This committee has under advisement for early publication a classification of accidents by cause and nature.

E. W. KOPF.

MEDICAL STATISTICS IN ENGLAND.

The organization of the Medical Research Committee in connection with National Health Insurance in the United Kingdom is, perhaps, the most important enterprise in British medical statistics since the establishment of the statistical department of the British Army in 1858.

Under the National Insurance Act of 1911 a fund of £55,000 became available for the use of the Committee. Researches are conducted in several departments of the Committee. The one on statistics is directed by Dr. John Brownlee. Some of the papers prepared and published by the department in the first year of the Committee's work are:

"Investigations into the Periodicity of Infectious Diseases." *Public Health*, March, 1915.

"Four Studies in the Meaning and Relationships of Birth and Death Rates." *Journal of Hygiene*, Vol. XV, No. 1, July 30, 1915.

"Historical Note on Farr's Theory of Epidemics." *British Medical Journal*, Aug. 14, 1915.

After the beginning of the war the Committee offered their resources to the War Office. The statistical department is now conducting the compilation of the sickness statistics of the Home and Expeditionary forces. Arrangements for transcribing the records of the military hospitals have

Digitized by Google

been made. In addition, the statistical department has undertaken the sorting and classification of the medical and surgical case sheets from the hospitals. This work renders case history information readily available for the use of medical officers with the forces abroad.

Dr. Brownlee and his assistants have also made a statistical investigation into the principal characteristics of the male population of the German Empire and have transmitted the report to the British Government.

The first annual report of the Committee is now available.

E. W. KOPF.

Unmarried Girls with Sex Experience. Bureau for Social Research of the Seybert Institution, Philadelphia, Carol Aronovici, Ph.D., Director. Bulletin I. 48 pp.

This is a study in a small way of the immoral girls under institutional care in the city of Philadelphia. Altogether 616 girls were studied, 392 in the House of Correction, 147 in the House of Refuge, 55 in the Magdalen Home, and 22 in the Midnight Mission. The principal results of the study were to show that in a large majority of cases the girls came from families in which the relationship between parents was abnormal, or where one of the parents was away from home. That feeble-mindedness was a decisive factor in causing immorality was not proved. To place such girls in domestic service was to expose them to peculiar temptations. It was recommended that industrial employment rather than domestic service be found for girls discharged from these institutions.

The statistical evidence afforded by this study corroborates the evidence afforded by other investigations of a similar nature in this country. The tables as a rule are well planned and intelligible to the lay reader. Table II, showing for Sleighton Farms their first parole placements according to causes of return, is not as clear to the untrained reader as it might be. The advisability of including percentages in a table where the total number of cases is quite small seems rather debatable. Only 22 cases from the Midnight Mission were studied and yet the items making up this total of 22 are distributed according to percents. Where the total number of cases is less than 100 it is seldom desirable to include a percentage table.

This first Bulletin seems to promise an interesting series of studies from the Bureau for Social Research of the Seybert Institution.

WM. B. B.

Digitized by Google

NEW STATISTICAL BOOKS ADDED TO BOSTON PUBLIC LIBRARY DURING QUARTER ENDING MARCH 31, 1916.

PREPARED BY LIBRARY COMMITTEE: ROGER W. BABSON, EDMUND E. DAY, HENRY J. HARRIS.

Donald, W. J. A. The Canadian iron and steel industry. A study in the economic history of a protected industry. Boston. 1915. Maps. Tables. [Hart, Schaffner & Marx Prize essays.] Bibliography, pp. 359–366. 9338.471

Dunbar, Donald Earl. The tin-plate industry. Boston. 1915. Charts. Tables. [Hart, Schaffner & Marx Prize essays.] Bibliography, pp. 125–128. An economic study of a protected industry. 9338.419a2

Haig, Robert Murray. The exemption of improvements from taxation in Canada and the United States. A report prepared for the Committee on Taxation of the City of New York. New York. 1915. Plans. Tables. 9033.373

Hasse, Adelheid Rosalia. Index of economic material in documents of the states of the United States. New Jersey. 1789–1904. Washington. 1914. [Carnegie Institution of Washington. Publication.] 9016.3073a13

Hepburn, Alonzo Barton. A history of currency in the United States. New York. 1915. Tables. Bibliography, pp. 479–495. 9332.4973a3

Huebner, Grover Gerhard. Agricultural commerce: the organization of American commerce in agricultural commodities. New York. 1915. Maps. Tables. Facsimiles. Bibliographies at the end of most chapters. 9380.73a6

Huebner, Solomon S. Life insurance: a text book. New York. 1915. Tables. Bibliographies at the ends of some chapters. 9368.3a65

International Institute of Agriculture, Rome. Bureau of General Statistics. Bulletin of agricultural and commercial statistics. [Monthly.] Vol. 2 (no. 11, 12), 3 (no. 2, 6–12), 4, 5. 1911–14. 9338.1a23

Shriver, J. Alexis. Canned-goods trade in the Far East. Washington. 1915. Tables. 9382.73a35.92

—Pineapple-canning industry of the world. Washington. 1915. 9382.73a35.91

United States. Department of Agriculture. Weekly news letter. Vol. 1–3 (no. 1–17). Washington. 1913–15. Illus. Plans. Previous to vol. 2, no. 35, April 14, 1915, the title was, Weekly news letter to crop correspondents. 9338.173a8

—Bureau of Foreign and Domestic Commerce. Banking and credit in Argentina, Brazil, Chile, and Peru. Prepared by Edward N. Hurley. Washington. 1914. Table. 9382.73a35.90

Economics of Retailing. By Paul H. Nystrom. Ronald Press Co., 20 Vesey St., New York, 1915. 407 pp. $2.00, postpaid.

The significance of this book to statisticians is its demonstration of the serious lack of fundamental statistical data in this great field of business

Digitized by Google

activity. It is possible to learn from the census reports how many establishments, employees, etc., are engaged in manufacturing, mining, and agriculture. The value of the products and the plants can also be ascertained in the same way; but nobody knows how many stores or storekeepers there are in the country and nobody can find out how many people are employed or how much they earn in the aggregate. It is a matter of common knowledge that an increasing proportion of the incomes of the people of this country are being expended in retail stores, year by year. But neither the amount of the gross income of the people, nor the proportion so expended is known with even approximate accuracy.

There can be no doubt that the retail store is one of the most important links in the chain of activities and processes by which the wants of the consumers are met by a ceaseless flow of goods from producers. It would be interesting and enlightening to know just how much this service costs. The author states that the "distribution" of goods, which includes transportation, insurance, and storage as well as the selling, probably costs more than the manufacturing. It is generally recognized that there is great inefficiency and waste in retailing. Here then is the place where statistics should be made available, if we are to deal intelligently with the rising cost of living.

The author has made excellent use of what figures he could get while at the same time he points out their inaccuracies and inadequacy.

The work is what the title indicates, a study of the economics of retailing. It combines in a very effective way the descriptive and analytical methods of treatment. It fits in splendidly between general economics on the one hand and the handbooks and treatises of a more technical character in business management and salesmanship on the other. It supplements general economics where many students have felt the need of a fuller treatment, and it affords a historical background and a scientific groundwork for the art of the merchant and the salesman.

The book is well written and is presented in a satisfactory typography. It should be interesting and easy reading for both business men and student.

CARROLL W. DOTEN.

Massachusetts Institute of Technology.

RESOLUTION.

At a Meeting of the American Statistical Association, held in New York City, May 5, 1916, the following resolution was adopted by unanimous vote of the members present:

WHEREAS, The Joint Resolution introduced in Congress by Senator John D. Works, of California, making it unlawful for any officer or employee of the United States Public Health Service to be, or to become, a member or officer of, or in any way connected with, any medical or private health association or organization of any kind, if passed by Congress, would limit the usefulness of a very important branch of government service; and

Digitized by Google

WHEREAS, It would deprive the public health and statistical associations of the country of the benefit that accrues to them and to the public at large from the direct contact and association with the officers of the United States Public Health Service, who have contributed largely to the knowledge of disease and its prevention; and

WHEREAS, The efficiency of the government service is enhanced by the participation of its officers in the activities of scientific societies; and

WHEREAS, The principle underlying Senator Works' resolution, if applied in one branch of the government service, might be extended to other branches and impose regrettable limitations upon the usefulness of government officers pursuing scientific work,

Therefore, Be it Resolved, That the American Statistical Association hereby records its emphatic disapproval of Senator Works' Resolution; and

Be it Further Resolved, That copies of this resolution be sent to the members of the Committees of Congress to which this resolution was referred and that it be given publicity in the daily press of the country.

Attest:

CARROLL W. DOTEN,
Secretary.

COMMITTEES.

The following committees have been appointed since the Proceedings of the Annual Meeting were published:

Committee on Fellows:

Frederick L. Hoffman, term expires 1920.
E. Dana Durand, term expires 1919.
Wesley C. Mitchell, term expires 1918.
Charles H. Verrill, term expires 1917.
John L. Coulter, term expires 1916.

Nominating Committee:

William Z. Ripley, chairman, Fred C. Croxton, W. S. Gifford.

Members of Committee on Local Arrangements: (Annual Meeting, Columbus, Ohio.)

Fred C. Croxton, Mary Louise Marks, and Carl J. West.

Digitized by Google

AMERICAN
STATISTICAL ASSOCIATION

NEW SERIES, No. 115. SEPTEMBER, 1916.

SCOPE AND METHOD OF STATISTICS.

By Harald Westergaard, *Professor of Social Science in the University of Copenhagen.*

INTRODUCTION.

By Walter F. Willcox.

It is a pleasure to welcome to the pages of these *Publications* the writer of the following weighty article and to express my own debt to him, a debt which began almost a quarter of a century ago with the beginning of my work as a teacher of statistics and has steadily deepened with the passage of the years. I know of hardly any writer in this field from whom I have learned more and whose work I can commend to others more unreservedly. That I am not alone among Americans in this opinion is suggested by the fact that years ago Professor Westergaard was invited by the United States to represent the guild of European statists at the Congress of Arts and Sciences in St. Louis and to give there an address upon demography. As he was unable then to accept, the present article may be deemed his first appearance before an American public and it seems appropriate, therefore, for me to add a few personal words about his career.

The son of a distinguished Oriental scholar, he himself turned first to mathematics and later to the social sciences. In these departments his academic work was supplemented by extended journeys in Germany, Austria, Switzerland, France, and England, during which he studied social problems, statistics, and insurance. At the age of 28 he presented to the University of Copenhagen a work on mortality outlining

Digitized by Google

and appraising recent investigations of the death rate in various classes of the population. This brought to him the University's gold medal offered as a prize for the best treatise on that topic. His work was published in German under the title, *Principles of Mortality and Morbidity*, and at the same time the author, who had already been employed for two years in the Danish Insurance Office, became a professor of social science at the University which had thus crowned his work. His book is an effort to analyze the various causes or influences, such as age, sex, marital condition, environment (urban or rural), race, religion, occupation, and diet, which affect man's death rate and to measure, or at least to point out the sound method for measuring, the effect of each influence upon mortality. Nearly twenty years later it appeared in a new edition completely rewritten and greatly enriched by the results of the author's intervening studies and those of other scholars in the same subject. For years I have found it an indispensable repository of fact and above all of argument, critical and constructive, in the wide field it covers.

Professor Westergaard's second book, appearing after an interval of eight years, grew directly out of his teaching. It is an *Outline of the Theory of Statistics*, published simultaneously in Danish and German editions and designed especially for the use of students in social science at the University of Copenhagen. Perhaps no better introduction to the present article can be found than a few words from the preface to his second main work:

Writers on the Theory of Statistics are often satisfied with developing the methods employed in special fields, like mortality, merely adding a few general remarks about the regularity of statistical phenomena and its consistency with our ideas of human freedom. This blinks the main question; namely, Under what conditions and within what limits is the regularity of statistical phenomena, "the law of large numbers," manifested? To sidestep that question has advantages, because it relieves one from entering upon a mathematical analysis; but it has also serious disadvantages, because it deprives statistics of its foundation. In following that procedure the statist cannot decide whether he has a solid basis

Digitized by Google

for his work because he does not know whether the law of large numbers is exemplified in the material before him. In the first part of the following book I have set forth the fundamental principles of the calculus of probabilities and then undertaken a task often overlooked, that of investigating whether these principles are exemplified in the results of statistical observation.

This second work of Westergaard, like the first, has been thoroughly revised and rewritten within the quarter century since it was published and is now about to appear in Danish. I sincerely hope that it may appear also in an English edition.

The present article is an epitome of the author's mature conclusions on a subject upon which American statists need to profit from the results of European scholars. The author's familiarity with the best English experience, perhaps suggested by the fact that the first edition of each of his two main works is dedicated to an English scholar and insurance expert, appears on nearly every page. He is wonderfully well informed also regarding the development of his subject from the beginning and his chapters on the history of statistics and of mortality statistics are marked by a combination of erudition and analytical power which raises them high above almost any other work and makes them a model in that field. For American scholars this is the more important because hardly one even of our scholars' libraries possesses the books needed for independent investigations in the history of statistics and yet perhaps no study provides a better thread to guide our researches past the pitfalls which trapped many former investigators and into the fruitful fields of statistical analysis.

In another respect, also, Professor Westergaard has much to teach our American statists. Those persons, mainly government officials, who are engaged in prosecuting statistical inquiries are seldom inclined by interest and training to use subtle or unusual methods of refined analysis upon the figures they have produced and thus fail to extract their full meaning. But they often have a just and sometimes an exaggerated appreciation, developed by their own sad experience, of the many errors from which their results suffer and of the uncertainty or at least the considerable margin of error in their

Digitized by Google

conclusions. This seems to be true of nearly every official American statist since Walker and Wright. On the other hand, such American scholars, trained in higher mathematics, as apply refined and modern methods to statistical data are, in the main, university men whose experience has seldom given them an opportunity to estimate the margin of error in official returns. They frequently accept without question figures so wide of the truth as to be most unsafe. For example, my own experience, including membership on a committee appointed to probe the divergencies between the figures of the Census Office and those of the Department of Agriculture, has made me sceptical of the annual returns showing acreage and yield of the staple American crops, returns now published under the more descriptive title of crop estimates. For this reason I have not thought it safe to use them in any published discussions. And yet elaborate theoretical structures employing the most refined methods have been built on these figures apparently without inquiry as to whether the foundation is strong enough to support the superstructure. These two classes of investigators, the producers of statistical data and the analysts of the results, need a closer coöperation and a more intelligent sympathy with each other's work.

It is noteworthy that Professor Westergaard, whose contributions to the refinements of statistical method have been of the first importance, should close his present article with the suggestion that our first object "must be not so much to prepare refined statistical methods as to provide useful observations. . . . At present we are more in need of statistical data than of theoretical investigations. . . . We want statistical observations covering new fields and here an enormous amount of work remains to be done" (page 52). In the United States both types of work are sorely needed, but more important than either alone is a sympathetic comprehension by each group of the aims and the difficulties of the other. In this way, and in this way alone, the new material needed can be most rapidly secured and the new methods applied only to significant material. Such a lesson of open-eyed and sympathetic coöperation Professor Westergaard is preëminently qualified to teach.

Digitized by Google

SCOPE AND METHOD OF STATISTICS.

By Harald Westergaard.

I.

It might appear absurd to discuss the scope and method of a science with at least 250 years of history, an evolution from the trifling beginning as "political arithmetic" to its present position as an important, or rather an indispensable, instrument in the complicated machinery of modern government and an auxiliary science in many branches of human thought. It might be supposed that no unsettled questions about scope and method were left, that on the contrary the tradition of many years would enable the great army of statisticians to follow a definite plan, without any doubt whatever about their final goal.

Still it is a fact, that there are at present not one, but several corps of statisticians, each trying earnestly to promote the science, but hardly able to coöperate for lack of mutual sympathy and sometimes acting in direct opposition to one another.

I hope, therefore, that it will not appear entirely useless to try to throw some little light on this problem. In doing so I shall first give a brief outline of some essential points in the evolution of statistics.

In the seventeenth and eighteenth centuries we find three entirely different lines of evolution, viz., the calculus of probabilities, "political arithmetic," and the comparative description of states, the last discipline being the only one to which the name of statistics was appropriated, though this one had originally very little to do with statistics in the modern sense of the word.

Statistics as the comparative description of states was cultivated by Aristotle in his famous Constitutions of 158 states, of which only the Constitution of Athens has survived. To the same category belong the works of the Italian authors Sansovino and Botero, dating from the end of the sixteenth century, those of their contemporary in France, Etienne Pasquier, who wrote *Récherches de la France*, and those of Pierre d'Avity, who published his great work, *Les Estats du Monde*, a little later (1614).

Digitized by Google

It was, however, in Germany chiefly that this study was cultivated. The well-known polyhistor, Hermann Conring, lectured on it for many years in the seventeenth century, beginning his description with Spain and ending it with Japan, Morocco, and Abyssinia; and in the eighteenth century we meet in Germany a whole series of writers from Achenwall to Schlözer, dealing with this theme. Most of these works had a family resemblance containing descriptions of the several countries, their climate and population, their produce and constitution, etc., but with hardly any numerical data; for these German authors did not take much interest even in the scanty numerical material which was at hand. Sometimes these descriptions were printed in a schematic form, giving under each heading for each country a short verbal description; in this way the so-called "Tabellen-Statistik" arose.

These tabular descriptions might be looked upon as signs of degeneration; but, clumsy and naïve as they often were, they actually indicate progress. For gradually the authors of *Tabellen-Statistik* began to fill their paragraphs with numerical facts, as material of this kind accumulated. In particular comparative tables with facts relating to population were introduced. One of the most important contributions of this kind was written by Crome (1785), with tables containing the number of inhabitants, area, density of population, and the "possible number," that is to say, the population assuming an average density of 3,000 inhabitants per (German) square mile. Crome tried to criticize in a really scientific spirit all the figures at hand and to make the best possible estimates. Thus a bridge was built between this discipline and statistics in the modern sense of the word, these comparative descriptions of various countries being increasingly influenced by the growing abundance of numerical material. This process was not without its dramatic features. At the beginning of the nineteenth century an interesting discussion over the meaning of statistics took place in some learned German publications; the views of the antagonists seemed quite irreconcilable; each side maintained that its conception of statistics was the true one. It was maintained that the adherents of the *Tabellen-Statistik* simply "skeletonized" the noble science of statistics, that they were *Tabellenknechte*,

Digitized by Google

unimaginative "slaves of the tables." As in so many other scientific controversies, there was never a formal conclusion; by and by the discussion died away, and gradually the name *statistics* was transferred from this discipline to political arithmetic, as formally proposed in a masterly essay, written in the middle of the century by the famous German economist, K. Knies.

The foundation of the calculus of probabilities, the second line of evolution, was laid by Italian and French mathematicians, who were interested in the problems of games at dice, lotteries, etc. This discipline reached a very high development in the eighteenth century, culminating at the beginning of the nineteenth century in Laplace's masterly *Théorie analytique des probabilités* (1812).

The chief problem of the calculus of probabilities was discussed in J. Bernoulli's posthumous work, *Ars conjectandi* (1713) and further developed by several brilliant mathematicians. According to Bernoulli's theorem, the probability of a certain deviation from a given type will depend on the number of observations, so that by increasing the number of observations it is possible to get as narrow limits of deviation as we like. This is the famous "law of large numbers," which has had so prominent a part in many statistical discussions. In fact, if it is possible to calculate the limits of deviation from the standard, we can easily control statistical conclusions. If the average rate of infant mortality is, say, 0.20, and if in a group of 10,000 infants a rate of 0.21 is very rarely reached, in a group of 1,000,000 infants a rate of 0.201 will be equally seldom found and if there are 100 millions of observations, the limit is reduced to 0.2001.

Although the mathematicians clearly understood the significance of Bernoulli's theorem, they failed to exhaust it fully. Most of the mathematicians who developed the calculus of probabilities seemed to suppose, as an axiom which needed no test, that its theorems would hold good everywhere. Those theorems can in fact be applied to a goodly number of cases, not the least in games at dice, lotteries, etc; but the important question was left untouched, whether and under what conditions these theorems might be applied to the more complicated problems in vital or economic statistics. No in-

Digitized by Google

vestigation was made to test the applicability of the theorems in this field; hardly even was an attempt made with regard to the original field of the calculus of probabilities, in spite of the severe but rather unjust criticism by d'Alembert, who directly appealed to mathematicians to make experiments with regard to games in order to find the real laws of chance. Many questions of this kind still remain unsolved.

The third line of evolution began in England. The cradle of political arithmetic is to be found in London, where a merchant, John Graunt, in 1662 wrote his remarkable *Natural and Political Observations upon the Bills of Mortality.* This treatise is a keen and ingenious attempt to calculate a life table from the London bills of mortality, on the basis of the returns of cause of death only, no statement of age at death being made. Graunt's results probably were wide of the truth; thus he supposed that three eighths of the persons surviving at six years would die before reaching sixteen, and again, that the same proportion would die between sixteen and twenty-six. Probably the actual death rate was only a fraction of this estimate. But still we must admire Graunt's acumen and his grasp of these highly important problems, which in the following generations were to be studied by a number of keen investigators, including the brilliant Halley and Lavoisier.

Halley's Breslau table is the next step in advance, a real seven-league stride, hardly appreciated by his contemporaries. His paper, published in 1693, contained *in nuce* a whole theory of vital statistics and life insurance. And not only did he grasp the problems thoroughly, but there is also reason to believe that his numerical results were fairly correct, in spite of the imperfections of his data. Thus, the total number of inhabitants which he calculated seems to have been fairly close to the truth, and so, too, his life table; it gave a good picture of the chances of death in a non-epidemic period in those times, though mortality in years of plague might, indeed, rise to a much higher level.

In the following century several mortality tables were constructed, often based on much more complete observations than were at Halley's disposal; thus by Struyck and Kersse-

Digitized by Google

boom in Holland and Deparcieux in France, the latter for instance calculating most interesting life tables for monks and nuns. Not less important was the progress in Sweden (1748), where a whole system of vital statistics was founded, with a population register in each parish and complete lists of births and deaths. The first reliable mortality tables for a whole country were calculated by Wargentin from these data.

How rich the literature of political arithmetic had grown in the course of a century appears from Süssmilch's compilation in his great work, *Die göttliche Ordnung* (2nd ed., 1761). His work contained, moreover, several original contributions, including a life table which enjoyed a great reputation in spite of conspicuous defects, obvious even to his contemporaries.

Halley's contemporaries, King and Davenant, had chiefly economic statistics in view. We owe to Davenant a first attempt at a theory of political arithmetic. (*Of the Use of Political Arithmetic, in all considerations about the revenues and trade*, 1698.) A single quotation will suffice: "As for example, when the number of inhabitants in England is known, by considering the extent of the French territory, their way of living, and their soil, and by comparing both places, and by other circumstances, a near guess may be made how many people France may probably contain." Statistical conclusions were very often drawn in this way in those days. Unfortunately the keystone was wanting; Davenant failed to explain how to avoid erroneous conclusions or how to find the limits of probable error.

In France several renowned authors of the eighteenth century grappled with the problem of calculating the number of inhabitants. So Messance, who, in his *Récherches sur la population* (1766), found a population of 24,000,000; Moheau arrived at a similar result (1778). Their method is easily understood. Moheau finds, for instance, in representative districts, that each year on an average there would be one birth to every 25½ inhabitants; the average number of births in France being about 929,000, he estimates the population at 23,700,000. The number of deaths are in Moheau's opinion less reliable as a basis for estimate, though in the

Digitized by Google

present case, by using the deaths, he was able to reach nearly the same result.

Lavoisier made a step forward in political arithmetic (1791); taking Moheau's method of estimate from representative districts, he applied it to agricultural statistics, and calculated on the basis of representative observations the number of ploughs and of horses and cattle, as well as the harvest and the cultivated area of the country.

Ingenious as these calculations often were, they were somewhat unreliable and arbitrary, nevertheless, for it seemed hardly possible to check the results. Laplace pointed in the right direction, basing his calculations of the number of inhabitants on figures of population and births in selected districts in 30 departments. His estimate of the number of inhabitants was probably nearer the truth than the previous estimates. But even he did not reach the bottom of the problem; he did not prove that these observations were grouped around a typical value according to the calculus of probabilities. After Laplace not much was done to develop the theory of representative statistics or of indirect methods of finding the numbers required. This is probably connected with the fact that direct methods of complete enumeration, such as a census covering an entire country, gradually became popular in spite of the great distrust with which at first they were regarded. In the earlier period of statistics this distrust was very characteristic. Davenant preferred the books of the Hearth Office; Messance would even rely on statistics of the yearly consumption of goods as a basis for estimating the number of inhabitants. Even as late as the middle of the nineteenth century, there was in France and Belgium a lively discussion of the advisability of using census results in calculating mortality tables, several well-known statisticians trying to do without the census.

The census held its ground, however, and no modern statistician would doubt the general reliability of enumerations of this kind. Only in calculating the infant mortality most statisticians still prefer the births to the census results, but even here a tendency toward using census figures is perceptible.

Digitized by Google

At the beginning of the nineteenth century the outlook for statistics was rather favorable. Statistical offices were established in France and Prussia, a census was taken in the United States in 1790 and again ten years later. But in Europe the long war and the political reaction after the conclusion of peace delayed progress, so that the new era could not begin until about 1830. From that time on there has been continuous and rapid progress. The enthusiasm of those early years was astonishing: statistical societies were founded; journals like *Annales d'hygiène publique* or the *Journal of the Statistical Society* of London published numerous statistical articles; statistical offices were founded or reorganized. This enthusiasm has waned a little, but nothing has prevented a steady progress, an immense accumulation of numerical data, of reports, *enquêtes*, etc. Under this rapid evolution statistics could not be altogether free from dilettanteism; it is even surprising that the number of such untrained workers has been so small.

Quetelet (1796–1874) may be considered as the typical statistician of his period. His enthusiasm, his readiness to draw conclusions from a rather meagre collection of material, his faith in statistics as the queen among all sciences, his theory of the great regularity in all statistical phenomena—all these features were common to many statisticians of his time, though they lacked the impress of his brilliant genius and his wonderful style. It is easy now to see his defects: Quetelet was too willing to believe in the constancy of types; he was fully convinced, for instance, that mortality did not change, though, even while he was writing, a considerable decrease of mortality was beginning. He saw too much in his "*homme moyen*," that remarkable being, who should unite all the qualities of the nation, as a representative of beauty and harmony; he was too optimistic with regard to international statistics and the usefulness of congresses. All this prevented him from seeing the variety which prevails everywhere, the constant changes in all conditions of human life. These mistakes were not peculiar to him. Mathematicians sought to find a formula by which mortality might be expressed as a function of age, a physiological law of mortality, so to speak.

Digitized by Google

The regularity of all statistical data being an axiom, it did not occur to the statisticians of the period that a thorough investigation into the conditions and the limits of this regularity was necessary.

But these erroneous ideas have to a great extent disappeared. In the following generation we perceive a certain reaction, a suspicion of laws and types. The statisticians mostly confined themselves to collecting exact details, giving, so to speak, a photograph of the conditions, but taking less interest in the causes behind all the phenomena. The doctrines of the present Italian school of anthropologists (Lombroso and his disciples) are perhaps the most closely akin to those held by Quetelet, but their views no doubt will gradually be changed under the influence of the strong and sober reaction which prevails among the statisticians of the present day.

As remarked above, the calculus of probabilities is one of the main sources of the theory of statistics. This discipline in the past century has had a healthy growth, some of the finest mathematicians having contributed to its development. It is sufficient to name Gauss (method of least squares) and Poisson (*Récherches sur la probabilité des jugements*, 1837). For many years statistics was too little influenced by these contributions. The official statisticians as a rule did not understand or appreciate the law of error. They were content to deal with statistical data in quite an elementary way with the result that much labor was wasted. It was often thought necessary to accumulate immense masses of data in order to prove results which might have been obtained from much less material, if the simple formulas of the calculus of probabilities had been used.

But here, too, came a reaction. One of the leaders was Woolhouse, who published his remarkable article, *On the Philosophy of Statistics* in 1872. Later came the brilliant contributions of F. Y. Edgeworth and of Karl Pearson's school, which took up problems of great significance concerning the physical and moral development of men, particularly with regard to heredity. Important contributions to the theory of statistics were made by the well-known German economist, Lexis, and his disciple, Bortkiewicz. Lexis took up the prob-

Digitized by Google

lem of regularity in statistical data, dealing principally with the ratio of males to females among new-born children.

A prominent feature of modern statistics is the immense development of the field. The progress of labor statistics during the last few decades is an instance of such extension of official statistical work. A further broadening of the field is due to the private efforts of pioneers, supported by the generous gifts of men like Francis Galton, and to the activities of life insurance and accident insurance companies and institutions investigating the mortality of the insured, or the frequency of invalidity, illness, or accident.

In this extension of the field of statistics much has depended on the attitude of the people towards statistical inquiries. The spread of education through good public schools has been an important factor in allaying popular suspicion. The former distrust lest statistical inquiries might prove to be a preliminary to increased taxes has changed into a general confidence and a willingness to coöperate with the authorities. Consequently in the future there will be rather an *embarras de richesses;* the difficulty will be not so much in gathering material as in mastering it, in digesting all these masses of reports which have been stored in the archives and on the book-shelves of statistical offices.

II.

What is the present outlook for statistics? Some hints of future possibilities are suggested by the foregoing outline of its development.

Let us consider first the future extension of the field of statistics. I have already mentioned the immense progress which has been made of late. It might perhaps seem difficult to make further advances, but the movement has been so rapid, that we can hardly think it will at once come to a standstill, and at all events some opportunities of further progress are evident. We need only to follow the lead of the old political arithmeticians. An essential feature in those early days was the frequent use of representative statistics, where observations covering the whole ground were not available. This method was frequently used too naïvely, but its principle

Digitized by Google

was not bad. Laplace pointed in the right direction, claiming that the representative districts should be very numerous and spread over the whole area. What remains to be done, is simply to develop a theory of these representative enumerations, stating the limits of the deviations from the true proportions and showing how to approach as nearly as possible to the truth. In fact in many cases it will be practically impossible to do without representative statistics. A typical instance is the production of milk in a country. No statistician would think of measuring the whole quantity of milk, yielded by all the cows in the country. It is only necessary to select a certain number of cows, fairly representative of the whole, and to test the quantity of milk from these cows on certain days. From these observations the whole product may be approximately estimated. So also with regard to the production of butter, meat, etc. The statistics of crops, too, will always depend, somehow or other, on representative data. Again, we may ask what is the amount of timber, which can be produced in a forest: we can only solve the problem by choosing certain select areas in the forest and proceeding to measure the timber in each. The rates to be charged by insurance offices is another example. By taking samples of workshops, cotton mills, theaters or farm buildings, we can find their characteristic fire risk, and from these observations we can derive the rates for similar buildings. Interesting examples of representative statistics are furnished by the Norwegian investigations into the distribution of incomes. By selecting certain places, certain streets and houses, and inquiring about the income and other circumstances of each person there, more detailed information could be secured than from returns covering a whole community, though in fact the modern systems of income taxes have made such knowledge easier to get than it used to be.

Related to the method of representative statistics are certain methods of interpolation. Supposing that the distribution of a certain population according to years of age is known, but that in dividing the population into groups according to occupation it was found too expensive, in spite of the enormous technical progress of recent decades,—

Digitized by Google

electric counting machines, etc.,—to give the same details, so that the occupation classes are divided only into broad age-groups, then we can fairly well calculate the detailed age distribution, on the basis of the age distribution in the whole population. The formula of interpolation applied may be a rather crude one, as was the case with Süssmilch's estimate of infant mortality in one class of population on the basis of that of another, or we may invent more refined methods; at all events the principle will be the same. As an instance we may take the comparative mortality among widowers and and married men. If the numbers living are classified in ten-year age groups, 55–64, 65–74, etc., it is clear that in each group the mean age of widowers will be higher than the mean age of married men. For this reason there will be proportionally more deaths of widowers, even if their mortality is just the same, and we will be tempted to draw erroneous conclusions from the observations. By means of interpolation we can reduce the age periods and thus remove the described source of error.

III.

Another group of most important problems relates to the correctness of the numbers observed.

Statistical observations are undoubtedly much more correct nowadays than formerly. But of course these observations are not and cannot be entirely accurate. In enumerating a population at a given moment we can get fairly close to the truth, but it will be hardly possible to register all the inhabitants without any omissions (for instance tramps) or altogether to avoid double counting. There will always be some little possibility of error left, and the question is, whether the limits of such errors and the effect they may have on our conclusions can be determined. Here we can learn from the old political arithmetic, for these authors, naïve and unpractical as they often were, frequently had a clear understanding of the importance of finding the limits of error in the statistical material. Süssmilch gives a good instance with regard to Berlin, Heysham with regard to Carlisle, and Milne's correspondence with Heysham concerning the Carlisle observations testifies to his anxious endeavor to

Digitized by Google

correct his materials as much as possible before calculating his mortality table. This ought to be a characteristic of every statistical report; the degree of accuracy of the observations should always be carefully considered. It would be a useful task to gather observations on the limits of error out of the enormous literature of statistical reports, and to try to systematize the results. Such a work would probably contribute not a little to our confidence in numerical data and would help us to draw valid conclusions in spite of the defects of the materials. There are in fact already available not a few discussions of this kind and undoubtedly more will appear as soon as the question has attracted the attention of official statisticians.

In several cases we have ways of testing statistical data. Sometimes certain questions are answered from two different points of view: for instance, in labor statistics, when reports are secured both from employers and from trade unions. In statistics of international trade the import of goods from one country into another should correspond to the export of the same goods from the former country into the latter.

Frequently it will not be possible without extra effort to determine the margin of error. For example, in age returns at some periods of life there is a temptation to overstate age: an infant 11 months old may perhaps be registered as having completed the first year of life; in extreme old age many persons, proud of their age, are tempted to exaggerate it a little. Again the gentler sex is often suspected of understating age. Further, there is the concentration on multiples of five or ten, common in census reports even where not only the age in years but also the date of birth is entered in the schedules; for many persons, ignorant of their exact age, are inclined to state it in years as a multiple of five or ten and then compute the year of birth by subtraction from the census year. A revision of the census figures is therefore necessary. In some cases a complete revision is possible. Thus in Norway, the country of centenarians, a careful revision of the data concerning persons in extreme old age was undertaken by comparison with the parish registers of births. In other cases we must confine ourselves to making a thorough investigation

Digitized by Google

of representative or select parts of the material, for instance certain parishes; from these parts it is often possible to draw valid conclusions applicable to the whole.

A certain pessimism is often encountered among official statisticians or economists who attempt to draw conclusions from statistics. It is maintained that the inaccuracies are often so great that it is impossible to get any reliable results. Here, then, is another important problem in the theory of statistics, viz., to determine the significance of the inaccuracies, to state to what extent it is possible to draw conclusions from statistical data, in spite of their imperfections. Here it would be most useful to form a theory of the applicability of imperfect data. For it can easily be shown that even extremely incorrect data may under certain circumstances allow of perfectly safe conclusions. It will be sufficient to discuss a few examples.

Let us suppose that we want to find the mortality of hospital patients suffering from pneumonia, grouped according to our estimate of their temperance or intemperance in the use of alcoholic beverages. Here the division line is very vague. Two persons with just the same habits may be registered, one as temperate, one as intemperate. The pessimists would maintain that observations of this kind are wholly without statistical value. But there is no reason to think that the observers will knowingly falsify the facts. There are patients whom every observer will judge intemperate and others who are as clearly temperate; between these two groups there are some whom it may be difficult to assign definitely to either group, and here different observers might make different decisions. Now let us further suppose that persons with intemperate habits suffering from pneumonia have a death rate of 40 per cent., whereas patients whose consumption of spirits is under a certain limit have a mortality of only 30 per cent., then as a consequence of the error in classification, the results may not show the true difference in relative mortality of the two groups, but it can easily be shown that the results will always indicate in which group the mortality is higher. If out of 1,000 patients recorded as drunkards only 900 are really such, whereas 100 ought to have been registered as temperate, and

Digitized by Google

a similar group of 1,000 "temperate" patients contains 100 who are really intemperate, then the observed number of deaths in the former group would be, not 400, but only 360+30=390, whereas in the latter group mortality would be 270+40=310 instead of 300. The difference, still marked, is merely somewhat reduced. If mistakes were still more frequent, so that, for instance, 400 of one group ought to have been registered in the other one and *vice versa*, then mortality would have been 240+120=360 and 180+160=340; there would still be a difference left. Only if more were reported incorrectly than correctly, would we get an erroneous result; where this is not the case, we find a difference testifying to an unfavorable effect of intemperance, even if we cannot measure this effect; we can only say that the difference is at all events not smaller than that which we have observed. But in statistics it often matters little whether we can measure the true effect of the causes which are acting or not, if we can only prove that such causes are in existence.

Or how does the health of children in parish-schools vary according to the economic circumstances of the parents? It may be that one observer will consider as belonging to a poor home a child whom another would report as well-to-do. Thus we have some chance of error in the observations, a fact, however, which does not prevent us from finding a difference in health among the two groups, if poverty has a deteriorating effect on health.

Or we may take a typical text-book example. It is a well-known fact that the death rates of legitimate and illegitimate infants at different ages apparently follow different laws. Up to a certain age, legitimate children have a lower rate of mortality than illegitimate, but after that age, a higher rate. In former days statisticians were often inclined to ascribe this phenomenon to a selection among illegitimate children, the more delicate being overtaken by an early death, whereas the healthier children survived. Thus the stock gradually grew healthier, whereas among children born in wedlock the opposite would be the case, delicate children being kept artificially alive, till at last they succumbed in spite of all the care of their parents. But there is an evident source of error,

Digitized by Google

viz., the legitimation of children by the subsequent marriage, of their parents. If such a child dies after the marriage, it will generally be registered as legitimate, though at birth it was reported as illegitimate. Consequently the fraction measuring the mortality of the illegitimate children is too small, of the other group, too large. As long as the mortality among legitimate children is smaller than among illegitimate, we know that there is an unfavorable cause acting on the latter and that the effect is even greater than the observed difference; but when the difference is negative, we are unable without further data to say whether the difference should be ascribed to defects in the data or to a selective process. We owe to the German statistician, Boeckh, a series of excellent investigations on this subject, proving that legitimate children in Berlin at no age of life had a higher mortality then illegitimate children.

This example is another illustration of the fact that we are often unable to measure the exact numerical effect of a cause, though we can prove that it really exerts some influence. But often it is unnecessary to know more than that. If we have found that a certain cause is detrimental to health, we may leave it to social legislation or to hygiene to bring about a change if possible, to secure which it is not so important to know the exact statistical quantities.

IV.

A third problem of wide import relates to the "law of error." We have seen that the old political arithmeticians did not realize the necessity of investigating the limits within which the observed values deviate from the standard, and, notwithstanding several efforts in this field of late years, much work has yet to be done. Outside of anthropometry, which presents interesting analogies to the law of error, not much had been done until Lexis took up the question, as mentioned above; Bortkiewicz made interesting investigations, particularly with regard to small numbers of observations (*Das Gesetz der kleinen Zahlen*, 1898); and I may refer to my own investigations beginning as far back as 1884.

Digitized by Google

Thus, in trying to explain the differences in the mortality, of married men and widowers, we might ask whether by chance there had been an uncommonly large number of widowers with ill-health, or living under such circumstances that they were more liable to attacks from epidemic diseases, or perhaps with a very small income, etc.,—circumstances, which have no connection with conditions of married or unmarried life. If this were the case, another investigation might give an opposite result and show a greater mortality among married men than among widowers.

In order to find the limits of these chance deviations, we may first consult the theory of the calculus of probabilities, asking whether Bernoulli's theorem will give us a point of departure. Where the theorems of the calculus of probabilities hold good, the observations will be grouped around the standard value according to the binomial law, found by expanding $(p+q)^n$, where p stands for the frequency of the event, q for its complement $(p+q=1)$ and where n is the total number of cases. Where n is sufficiently large, we can approximately measure the deviations from the standard by the "mean error" $(\sqrt{npq})$. In two cases out of three the deviation from the standard will be smaller than the mean error and in 19 cases out of 20 smaller than twice this quantity, and a deviation over four times the mean error will be exceedingly rare. Simple as this formula is, we can simplify it still more, if the value of p or q is not very great. If $p=0.1$ and $n=10,000$, we have a mean error, $\sqrt{10,000\times0.1\times0.9}=30$. If we disregard $q=0.9$ in the formula, the mean error will be only a little more; namely, $\sqrt{1,000}=31.6$; we have then only to take the square root of the number of deaths (np) as a standard of the deviations. Practically this change makes little difference, and where the value of p is still smaller, the difference will often be very small. Thus, if $p=0.01$, $n=1,000,000$, we have for the mean error $\sqrt{990}$ or 31.5, but leaving out $q=0.99$, we get $\sqrt{1,000}=31.6$, or a quite insignificant difference.

The use of this formula is evident. If the observed number of deaths is 1,020 or 1,030 instead of 1,000, we may look upon it as an ordinary deviation, which need not be ascribed to peculiar causes. But should the deaths increase for instance

Digitized by Google

to 1,100 or 1,200, we should naturally infer that some peculiar causes have been acting, and the next step will be to make further observations in order to get a clearer understanding of the matter.

The same formula in a little more complicated form can be applied to the chief problem in medical statistics; viz., to find whether a particular method of treatment of a disease is effective. Let the mortality of patients suffering from the disease be p_2, when treated with a serum, p_1, when treated without it, and let the numbers of observations in each case be n_2 and n_1. Then the mean error of the difference between the frequencies of dying in the two groups will be $\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}$, and we can get an approximation by putting the observed relative values instead of p_1 and p_2. If, for instance, 200 patients have been treated in each group, the numbers of deaths being 40 and 80, the observed difference thus being 0.20, we shall have the mean error of the difference $\sqrt{0.008+0.0012}=0.045$. The observed difference is four to five times the mean error, and we are consequently justified in believing in the favorable effect of the treatment by serum.

But the question arises: is this test applicable, or do the casual deviations follow some other law? As mentioned above, the students of the calculus of probabilities were often mistaken, for explicitly or implicitly they considered the theorems as axioms, which would hold good at least in all games of chance. In fact this can only be tested by experience. Although d'Alembert in his controversy with other mathematicians had the worst of it, he was right in wishing observations; he had actually touched a weak spot in the theory.

It may, however, be considered as sufficiently proved by experience that the calculus of probabilities can be applied to the problems of games and similar questions. How shall we explain this result? If we toss a coin in the air, or throw dice on a table, there will always be numerous circumstances which we cannot register at all, but which combined contribute to the definite result of the games. The die is on the table in a certain position and at a certain distance from the player; the position of his hand varies; the die is thrown with varying

Digitized by Google

force; etc. If all the circumstances each time were exactly alike, then the same thing would always happen; if the die had once shown a six, the same number would be shown at every throw; if one card had once been drawn, the same card would constantly appear, etc. In fact all these innumerable individual causes coöperate, so to speak, so that their effect partially disappears, and we are consequently enabled, within the limits given by the "law of error," to foretell the collective result of a number of experiments. We cannot tell what number the die will show at a single throw; but if we repeat the throw, for instance 12,000 times, we know from experience that about 2,000 of these cases will give sixes, and we can with the help of the "mean error" easily find the limits of the possible deviation from this standard. The mean error is 41, consequently we will very rarely expect a deviation of, say 120 or 160; if the die should give sixes 2,200 times instead of 2,000, we should certainly look upon it as "false"; it must have had some quality favoring this event, possibly a certain part of it was heavier than a corresponding part of other dice. If, then, we have found sixes 2,200 times, we know that there is a peculiar cause which we shall have to find in some way or other. The greater the difference, the easier will it be to find the cause.

Now, in vital or economic statistics we have an analogy to these phenomena. Here, too, a certain regularity prevails; particular individual causes are, so to speak, eliminated in the collective results. No two persons are alike; one will be attacked by an epidemic disease, another will escape; one will marry, another remain single; etc., but on the whole in a certain class a certain number of deaths or marriages will take place. The question now is, What are the limits of deviation from the average? Can we find a conformity to the binomial law of error similar to that found in playing at dice or cards, or do other laws of frequency apply to these phenomena?

In vital or economic statistics most numbers have a much wider margin of deviation than is experienced in games. Thus the death rate, the birth rate, the marriage rate, or the relative frequency of suicide fluctuates within wide limits. But it can be proved that, by dividing the observations, sooner or later a marked tendency to the binomial law is revealed in

Digitized by Google

some part of the observations. Thus, the birth rate varies greatly from year to year; but every year nearly the same ratio between boys and girls, and the same proportions of stillbirths, and of twins are observed. The causes which bring about an increase or decrease in the number of births thus have apparently the same influence on both sexes.

A curious illustration of this is found in the statistics of divorces in Berlin, 1899–1908.

Year.	Number of Divorces.			Percentage of Divorces.	
	Total.	Protestant.	Mixed Religious Belief.	Protestant.	Mixed.
1899	1,608	1,261	232	78	14
1900	936	732	128	78	14
1901	984	769	149	78	15
1902	1,227	972	174	79	14
1903	1,269	981	198	77	16
1904	1,376	1,048	229	76	17
1905	1,421	1,103	219	78	15
1906	1,639	1,271	266	78	16
1907	1,781	1,350	298	76	17
1908	1,868	1,444	291	77	16
	14,109	10,931	2,184	77	15

It appears from this table that the absolute number of divorces in Berlin has fluctuated widely. A change in the divorce law accounts for the low number in 1900 compared with that of 1899, but except for this the numbers have been constantly on the increase. The distribution according to religious belief, however, is very regular; calculating the mean error of the percentage numbers, we find that the fluctuations are within the limits of the binomial law. Hence we may reasonably conclude that the causes affecting the numbers of divorces act with the same force on the two groups of marriages; the individual causes affecting the percentages in each group have collective results of the same nature as in the case of games.

It is easy to multiply illustrations of this kind, showing the grouping of relative numbers about a standard. By the test of experience, then, it is possible to apply the binomial law of frequency in many fields of vital statistics. It seems more difficult to apply the law to absolute numbers, calculating for instance, the number of births or deaths in a given year,

Digitized by Google

for the fluctuations of these numbers generally exceed the limits of the binomial law. Still we can sometimes find the causes of variations and thereby calculate the absolute numbers more precisely. As regards the number of births it is necessary to make allowance for a tendency to a decrease of fertility; the number of legitimate births depends on the number of married couples, the age of the parties, and the duration of their marriage. A calculation for Denmark through a period when there was no noticeable tendency to a decline of fertility, taking into consideration only the fertility of married women by five-year age groups, in urban and rural districts, but without making allowances for duration of marriage or age of husbands, gave a favorable result, showing that it was not impossible to calculate with tolerable accuracy the number of legitimate births in a particular year from the experience of previous years.

As to the deaths it seems more difficult to calculate the absolute numbers, because meteorological and economic causes exert an influence the numerical effect of which it is hardly possible to measure. By leaving out of consideration ages at which variations in climatic conditions cause great fluctuations in deaths (infancy), we may make some advance towards the binomial law. But there is no difficulty in getting several important results concerning relative numbers. The level of mortality may be very different from year to year, but we can perceive a tendency to the binomial law in the relative numbers, the death rates by age, sex, occupation, etc. By observing the death rate of two occupations in the same period and in the same country we may find that mortality in these professions is rising or sinking with the general level, and we are in a position to conclude from the figures which occupation is more healthful.

So, too, in the case of divorces. The absolute number varies greatly, as we have seen, but the distribution according to religious belief is rather regular. Knowing that 3 per cent. of the dissolved marriages in Berlin are Catholic, whereas about 5 per cent. of the weddings are Catholic, we are warranted in inferring that Catholic marriages are less likely to be terminated by divorce than others.

Digitized by Google

In economic statistics it is more difficult to apply the binomial law, the conditions being as a rule much more complicated. But if we carefully single out homogeneous observations, we shall often find a decided tendency towards the binomial law. I may refer to interesting examples in H. L. Moore's work, *Laws of Wages* (1911). Another instance can be taken from the Copenhagen statistics of wages (1909). We have here the following distribution per 1,000:

Range of Yearly Income in Crowns.	Wage-earners per 1,000.	Range of Yearly Income in Crowns.	Wage-earners per 1,000.
Under 100	7	Under 900	416
100–200	24	900–1,000	18
200–300	37	1,000–1,200	48
300–400	40	1,200–1,400	139
400–500	52	1,400–1,600	167
500–600	62	1,600–1,800	119
600–700	70	1,800–2,000	77
700–800	66	Above 2,000	16
800–900	58	Total	1,000
Under 900	416		

This distribution has little resemblance to the binomial law: it has two maxima. But by dividing the observations into several classes we may discover a tendency towards the law. Thus, we can ask how each occupation is grouped around its special average, how many persons have, for instance, at least 100 or 200 crowns yearly more than this average, etc. If we then displace all these series of observations so that all the centers coincide, we shall find how all these working men and women together are grouped around their averages, and we may compare this to the exponential law (which closely resembles the binomial law). The result will be as follows:

Deviation from Average	Frequency.		Frequency Expected According to Exponential Law.
	Under the Average.	Above the Average.	
(Crowns)			
0–100	223	212	195
100–200	154	148	152
200–300	84	73	91
300–400	33	29	42
400–500	6	19	15
500–600	3	9	4
600–700	2	1	1
700–800	2	0	0
800–900	2	0	0
Total	509	491	500

Digitized by Google

On the whole it may be said that there is a tendency toward the exponential law; and this will hold true, if we deal separately with males and females, with skilled and unskilled workers. If some few occupations show conspicuous differences from the law, it always seems possible to find the reason. Thus, in gasworks there seem to be two distinct groups of laborers, each group with its own average.

In anthropometry we have a good analogy to the theorems of the calculus of probabilities. The distribution of height among young men registered for military training is an excellent example. This analogy enables us to find interesting results with regard to typical differences between rural and urban population, rich and poor, etc.

If no conformity to the binomial law can be found in a series of observations, the first task is the classification of the data. The most obvious classifications are by sex, age, profession, residence, etc.

If we do not at once reach the binomial law, we must try further subdivisions. Since each of these classes or subclasses has its peculiar conditions, and is under the influence of different systems of causes, we can say that the more quickly we reach the point where the binomial law holds good, the fewer causes have been acting, whereas the greater the deviation from the binomial law, the greater the number of active causes. An instructive instance of this is found in the discussion of the balance of male and female births—one of the most interesting chapters of the history of statistics. According to Hofacker and Sadler, if proportionally many young men died—as in the present war—then nature would try to restore the balance, so that the proportionate number of male births would increase; couples where the wives were much younger than their husbands would have a relatively numerous male issue. The difficulty here is simply that we generally find very soon a conformity to the binomial law. Moreover, statisticians who dealt with this problem frequently misunderstood the results; in fact very often they had not observations enough, the differences they found being within the limits of the law of error. The late Dr. Geissler found two sets of causes acting against each other and thus creating a tendency to balance:

Digitized by Google

there are families with relatively many chances of getting children of one sex only, but where this uniformity is once broken, then there is greater probability in future of getting children of the other sex.

V.

As far as I can see, the typical frequency curve in all vital, social, or economic statistics is always the binomial one; but it will require much investigation by statisticians to get to the bottom of this question, to prove whether this supposition is right, or under what conditions the observations will show a tendency to the binomial law. This third problem of the theory of statistics will require much patient labor in careful analysis of each separate class of material. If the observations show the validity of the binomial law, then we know that we can use the above described test of the mean error with little or no hesitation and that on the whole the calculus of probabilities is applicable.

Many statisticians have tried to apply other frequency curves than those with which I have here dealt. We are indebted especially to Karl Pearson and his disciples for important investigations on this question, leading, for instance, to curves with "skew variations," etc. The binomial law is generally a little askew, but for all practical purposes we may set aside this small difference from the symmetric form. Very often, however, we meet with distributions of quite another form; how are we to deal with such phenomena?

I may here insert that it is not as a rule necessary to know the mathematical formulas of these distributions. We may deal with them well enough, if we only know by experience the form of the curves. Distributions according to the binomial or the exponential law need no strict mathematical treatment, the observations from games will give us sufficient materials for finding the form of the curves rather accurately, even for more complicated data. A curious instance is presented by the Danish statistics of fertility. If marriages of 10–15 years' duration are classified according to social position, certain classes, represented by 3,759 families, are found where the birth rate is high, and others, represented by 2,082 families, where it has declined to a low figure. If the couples in both

Digitized by Google

groups are distributed according to size of family, after a certain size of family is reached, we find that the numbers in both groups show a close correspondence. Out of 2,082 families with low fertility, 192 had more than six children each, distributed in practically the same way as the corresponding 753 couples in the group with a high birth rate. But these 753 correspond to 3,006 couples in the group with the high birth rate with six children or less; in the group with a low birth rate we may separate out a corresponding number, 766 couples with six children or less,—this number bearing the same proportion to 192 as 3,006 to 753. Assuming that these 766 are distributed as in the larger group, we have the following figures:

Number of Children	958 Couples with a High Birth Rate		1,124 Couples with a Low Birth Rate		The Whole Group	
0	94		214		308	
1	60		180		240	
2	99		260		359	
3	108		187		295	
4	128		196		324	
5	149		55		204	
6	128		32		160	
Under six		766		1,124		1,890
7	102		..		102	
8	51		..		51	
9	25		..		25	
10 and over	14		..		14	
More than six		192		..		192
Total		958		1,124		2,082

We have thus found that 958 couples seem not to have limited the number of their offspring, whereas the remaining 1,124 couples appear to have had modern ideas. This is of course only a preliminary investigation. In reality we should have split the group into several classes, each with its own birth statistics. At all events it appears from these numbers, that the movement to limit the birth rate had reached more than half of the group concerned.

It is apparent in a case like this that it is unnecessary to find a mathematical expression for the distribution of families in the two groups. Nor will it be advisable. On account of the great rapidity of the decline of the birth rate the distribution

Digitized by Google

of the groups will undergo considerable changes, which cannot be foreseen in advance, at least as regards the group with a low birth rate. The curves which we should find would therefore not be typical, and it would be of no use to make the calculations.

But the same objection can be made to several other curves designed to express the distribution of observations in some field or other. One of the most interesting examples is that of the distribution of income among the inhabitants of a country. In trying to draw such a curve we shall certainly find a very skew distribution, the frequency curve differing immensely from the curve conforming to the binomial law. But if we try to divide society into different classes, we shall find in each one a separate distribution with its own center of gravity. In Danish income statistics the figures show a marked tendency for the separate groups to be distributed about an average in a way resembling the binomial distribution; but, if all the groups are taken together, the curve for the whole population assumes a "pyramidal" income distribution (Pareto's law). It is not safe to stop here; a simple change in the constitution of the society, a labor conflict resulting in higher wages, for instance, may give the "pyramid" another form. This will be the case, too, if the commercial classes get higher profits from their business, or if the share of rent decreases or increases, etc.; all these causes are to a certain degree connected with one another, but so much depends on unknown factors that we can not well foresee the changes of the figures. Taking each class alone, we have a much better foundation for our investigation, and then we can return to the total population to get a view of the collective results of all these peculiar movements. Using class statistics as a starting point, we are in fact much nearer the significant causes. If we seek a formula for the combined effects of all the causes in action, we run the risk of overlooking some, which it would really be exceedingly important to take into consideration. Much labor can indeed be lost in this way.

The number of sick days after an accident affords another instance. As there are always proportionally very few cases of protracted illness after an accident, the result will again be

Digitized by Google

a very skew variation or a pyramidal distribution. But instead of trying to give this distribution a mathematical expression it would pay better to classify accidents in special groups, each with it peculiar center of gravity.

The life table is another example. Here we have a combined expression necessary for practical use in life insurance, for calculating the present value of future assets and liabilities. But several mathematicians have erred in thinking that it would be possible to find a mathematical law of mortality, a physiological law, as it were. We have several formulas of this kind, by Lambert, Moser, Gompertz, Makeham. For a a certain period of life Makeham's formula is exceedingly practical, but after all it is only a beautiful formula of interpolation. It does not hold good at the end of life, for after the age of 80 mortality seems to vary according to another law, nor does it hold good in infancy. And from 20 to 80 the variation is by no means always as uniform as the formula represents; a new treatment of tuberculosis may cause a diminution of the rates of mortality in a certain period of life, the entrance into married life improves the health, military service causes an increase of mortality, etc. Under certain circumstances there will be even stronger influences, thus, for instance, the mortality of sailors, of persons with inherited diseases. A general mortality table will always be an aggregate of numerous special tables, and, if the groups which form the whole community. undergo change, then the life table itself will change, and it may prove impossible hereafter to apply the formula. For a close study of the causes at work it will be necessary first of all to take into consideration the mortality at each age and in each group.

VI.

Whether or not we use mathematical formulas for these combined expressions there are certain mathematical problems in dealing with these complex averages which a statistician must be able to master.

In trying to get comparable data it is often necessary to go through much preparatory labor. If the numerator and denominator in the fraction measuring the frequency of an event,

Digitized by Google

are not quite homogeneous, it may be necessary to make certain interpolations. We may have to find by interpolation from the census figures taken every tenth year the number of persons at a given age, who are exposed to the risk of death at a certain moment. On the whole we may here lay down the principle that this interpolation first of all should keep as near real life as possible, taking everything into consideration which might prove of importance, such as the fluctuations in emigration or immigration (as far as we can ascertain them), or of births and deaths. Complicated as these problems frequently are, it may be maintained that on the whole they have been solved with sufficiently close approximation. In fact, in many cases the different methods of interpolation lead to nearly identical results. This is true also with regard to adjustment or graduation, a process closely related to interpolation. An experienced mathematician as a rule will get nearly the same results whether he adjusts the figures graphically, "mechanically," or by some mathematical formula, with or without application of the method of least squares. The possible progress in this field will be therefore chiefly that of refinement, smoothing the values to make the changes appear as regular and even as possible. But on the whole, the theory of statistics cannot expect much improvement from such efforts beyond what has already been achieved.

The best means of treating the changes of a variable in a series of observations is the method used by Daniel Bernoulli in the eighteenth century in studying the influence of smallpox on mortality and further developed by Duvillard at the beginning of the nineteenth century. They used the fiction that all variations were continuous, so that it was possible to use the differential calculus instead of operating with finite differences. This method is of great value, as it simplifies many problems extremely; on the whole, it seems to have been generally adopted, though not always without hesitation.

According to this method we calculate, for instance, the force of mortality, a quantity measuring the risk of dying in the next infinitely short moment. This quantity being μ_x, where x is the age, we have as the number of deaths during a moment of length d_x, with l_x standing for the number exposed

Digitized by Google

to risk, $\mu_x l_x d_x$. If l_x signifies the number surviving at age x according to a life table, then the number surviving at age l_x+d_x will be $l_x-\frac{dl_x d_x}{d_x}$. The number of deaths during the moment d_x being also $\mu_x l_x d_x$, we have $\mu_x l_x=-\frac{dl_x}{d_x}$. If we are warranted in making certain simple suppositions with regard to the force of mortality, this equation will lead to simple expressions for l_x. But not only can we use the force of mortality as it stands, we can also split it into its elements, finding, for instance, the rate of mortality from a single cause of death, or a group of causes, and further asking what will be the result if one or more causes disappear. Or we may introduce another quantity, the rate of marriage, corresponding to the force of mortality, and ask how many persons will survive as single, how many will marry within a certain interval, how many married people will die in the same period, etc. It will cause no difficulty, if the force of mortality is supposed to be different among bachelors and married persons. The same methods can be applied to other problems, for instance, finding the number of persons who become invalids or who recover or who die after recovery. Here, too, we can use different values for the rate of mortality and allow for the influence of age and the duration of invalidity. Again, we can determine the probability of a person being sentenced for a crime, once, twice, etc. The problems are everywhere of the same kind and they present no particular theoretical difficulty. In fact, the main difficulty at present seems to consist in providing sufficiently correct and complete observations rather than in dealing with the materials afterwards.

VII.

Where are we at present with all these problems? We might describe the program of vital statistics as the task of following human beings from the cradle to the grave, taking account of all their chances, their good fortune and mishaps, their physical and intellectual growth, their criminality and moral defects, etc., etc. What will be a young man's choice

Digitized by Google

of profession? Will he follow in the steps of his father or not? What will be his chance of getting a better position in society or the risk of his being forced downwards in the struggle? What will be the danger of unemployment, and, after having lost a position, what is the chance of finding work again? What income will a man on an average earn in each occupation and at different ages, and what will be his economic position on reaching old age? Again, will education pay in the competition of life, will a boy who begins as apprentice in some trade get a substantial pecuniary advantage, or will he be actually no better off than an unskilled laborer?

It will be readily seen that there is still a long way to go. Thus, as to birth statistics, we shall have to add considerably to the present store of statistical data on fertility in different classes of society as influenced by the duration of marriage and the age of consorts. In marriage statistics we shall have to observe the different rate in different classes with allowance for the fact that very often a wedding is registered for two places of residence—that of the bride and of the bridegroom—and that the wedding frequently is followed by a change of occupation.

As to mortality statistics there used to be an enormous confusion. In the middle of the past century medical statistics suffered severely from dilettanteism, especially as regards the correlation between mortality and occupation. Much was accomplished in the way of clearing up the subject in England by the efforts of William Farr and his successors, but there is still a serious lack of material. The same may be said of statistics of sickness and invalidity, though observations during a century of experience in Friendly Societies are available and German compulsory insurance against invalidity, with its complicated machinery, has been in force for 25 years. Other problems have hardly been touched as yet, thus the mortality of persons suffering from certain diseases, such as syphilis, mental diseases, etc., the mortality of persons rejected by life insurance companies, the influence of heredity, etc.

Still less cultivated is the field of migration statistics. We have some data with regard to transatlantic migration, but not much more. Still it would not be impossible to find approxi-

Digitized by Google

mately the rate of migration for different ages (corresponding to rate of mortality). If we know from two consecutive censuses the numbers of persons born in a certain year, and if we know further the rate of mortality, we can find approximately the rate of migration (excess of emigration above immigration); and knowing in some places the numbers of those who immigrated a short time before the census, we have some data by which we can separate this rate of migration into its elements.

Still more complicated are the problems of economic statistics, for here we have frequently one more dimension. We have to deal not only with numbers or quantities, but also with their value in money. Only a very small part of the work has been done, and the scanty materials we have are frequently of doubtful value, because it has rarely been possible to separate them in such a way that the numbers are quite homogeneous. This is the main reason why statistics have so often been considered unreliable.

In agricultural and industrial statistics we have above all the problem of finding the quantity produced and its value in money; as mentioned above, representative investigations in this field are particularly useful or even indispensable. Other important problems to be investigated are the distribution of enterprises by size of farm or industrial establishment, and the working of the law of increasing returns in industry. Statistics of international commerce are somewhat developed, though there are evident defects in the returns, but of internal commerce we know practically nothing. The variations of price under the influence of supply and demand form another series of important statistical problems. Again what is the rapidity of circulation of a coin or a bank-note as it passes from hand to hand, and how can we get observations on the influence of bank reserves or of the rapidity of circulation on prices and the rate of interest? As a link between vital and economic statistics we have, moreover, the question as to the yearly income of a whole nation or a class of population, the statistics of public finances, shifting and incidence of taxation, etc. Even here only very few of the important problems have been thoroughly dealt with.

Digitized by Google

VIII.

It appears from this sketch of future problems, the list of which any economist might easily extend, that statisticians of the present day have no lack of work. But frequently we observe among statisticians, a certain despair, of their ability not in the collection but in the interpretation of statistical data. In consequence of this pessimism there are statisticians who confine themselves to giving a true description of the phenomena. But we must not forget that a true description cannot be given if we cannot find the relations of cause and effect that prevail, for else we run the risk of giving a confused mass of details instead of a clear outline of the main features of the observations concerned.

It must not be forgotten, however, that the task of the statistician is not so much to find the causality himself as to help others to find it. The statistician must be content if he can show that certain groups of numbers show marked differences, leaving it to physiology, meteorology, and other sciences to explain these differences. Thus his task is indeed a very modest one, but it is always satisfying to be able to put others on the right track or to spare them from some misapprehension of the facts.

Often enough the statisticians will be obliged to acknowledge that there are two or more causes which cannot be separated; it may even be that two causes have a much greater effect if combined than if working separately. So with regard to labor and capital. If a man works without any capital, his produce is quite insignificant; and, if capital is left alone, its product will be zero; no one can tell therefore which part of the produce is to be ascribed to labor and which to capital. A simple example will show how easily confusion arises. Suppose that certain goods to the value of $10,000,000 are exported. If now the exported quantity increases 50 per cent., and prices 20 per cent., then the value of the export will be increased to $18,000,000. If prices were unaltered, but the quantity only changed, then there would be an increase of value of $5,000,000; if prices only varied, we should have $2,000,000 more; thus we could account altogether for $7,000,-

Digitized by Google

000; the real addition was $8,000,000, or $1,000,000 more. In this case it is easy enough to see what the matter is, but very often the problems are much more complicated. An example from anthropometry will illustrate this.

Let us ask whether height is an inherited characteristic. To make the case as simple as possible, let us suppose that there are in a country two different types, each with its peculiar average, round which the individuals are grouped. A person belonging to the tall type may by chance be of small stature, and *vice versa;* and, if we divide the population into two groups according to height, we shall have a smaller or larger number of the short type in the group of tall men and *vice versa.* Supposing now that the type is inherited, whereas the casual causes of deviations from the typical height are not transmitted to the offspring, we have in the second generation two groups, each numbering persons of both types, but probably with an unequal distribution. Each type having its own center of gravity, the observed average in each group will lie between them, that is to say, the second generation will apparently show a smaller distance between the two sizes than was the case with the fathers. Galton and his disciples would speak here of regression, human beings having apparently a tendency to revert to former types, but it follows from the above that this regression is only a formal effect of the nature of the observations; it does not really lead us to an understanding of the causality.

IX.

Let us now suppose that we have a group of observations in which only one rate of frequency may be said to prevail; then, as shown above, we have the means of controlling our conclusions by finding the limits of deviation from the standard value. This is the most simple of the statistical problems. But generally the data are not homogeneous. Most frequently the material must be split up into several groups, each with its own frequency, and we have then the problem of getting an idea of all these relations together.

We can sometimes confine ourselves to studying each of the groups separately in comparison to other corresponding groups of observations. The two series may stand so clear of each

Digitized by Google

other that no doubt is possible. When in a certain occupation the mortality in each age group is decidedly greater than in the general population, it is clear that the occupation concerned is under the influence of some unfavorable cause. But frequently the differences are not all in the same direction, sometimes some of the rates in the occupation are lower than in the general population, and the evidence thus is apparently contradictory. This irregularity may perhaps be due to accidental causes, the influence of which would have been eliminated, had the number of observations been larger. It will therefore be necessary to combine these results in order to get a clear idea of the influence prevailing in the occupation. Here we have many methods of procedure. In determining which method is the best, it is a good principle always to keep as near as possible to the original observations, never leaving them out of sight. Taking, in the present case, the mortality of the general population as a standard, we may calculate some expression or other from the two series of observations; the mean duration of life, the value of a life annuity, etc. But, according to my experience, it is always best to choose a method which permits us to keep in view the original numbers. In this way it will be easier to avoid misapprehension and erroneous conclusions.

The simplest standard calculation in the present case is to find the budget of deaths in the occupation,—the deaths that would have occurred in the occupation if the mortality had been that of the general population or other group,—which for some reason we might consider typical. To illustrate this, I shall begin with a very simple case, that of barristers and solicitors according to English vital statistics (1900–1902). The following table gives all the necessary details:

Age Period.	General Population.	Barristers and Solicitors.			
	Death Rate per 1,000.	Population.	Actual Deaths.	Expected Deaths.	Death Rate per 1,000.
25–34	6.38	16,323	77	104	4.72
35–44	10.94	19,086	142	215	7.21
45–54	18.67	13,869	189	259	13.63
55–64	34.80	6,741	188	235	27.89
25–64	14.08	56,619	596	813	10 53

Digitized by Google

The simplest procedure would be to compare the rate of mortality for barristers and solicitors at all ages (10.53) with that of the general population (14.08), but as the age distribution is not quite the same, this comparison will not be without objection. The rates for each age, however, show very much the same proportion to the corresponding general rates, and we might stop with that; but, if we want a common expression, the budget of deaths will give a practical means of comparison. Under the supposition that the mortality is the same as in the general population, we find in the youngest group 104 expected against 77 actual deaths, in the next age group we have 215 against 142, etc., altogether 813 against 596, the actual mortality being thus only 73 per cent of the expected one. The mean error is 29 or about $^1/_7$ of the deviation (217); the mean error in the single age groups is between $^1/_3$ and $^1/_5$ of the corresponding deviations. The mean error is then proportionally much smaller, if we take the total, than if each group is taken individually. This is a great advantage, for often there are so many age groups (for instance, yearly intervals) that the deviations from the type are very irregular, whereas the total may allow of quite a safe deduction. The separate bits of evidence may be rather uncertain, but if combined they may warrant a definite conclusion.

The calculation of expected deaths will be particularly advisable in case the age distribution of the deaths is unknown. As remarked above, we cannot draw sound conclusions from the crude rate of mortality, but if the expected mortality is calculated, we have in the actual and the expected numbers, two homogeneous quantities, which we can safely compare. In a case like this it may be objected that mortality in one age group very well might be greater than the expected; we cannot tell whether mortality in the age group 25–34 actually has been lower or higher than in the general population. But at any rate it is clear that in some age group or other mortality must have been decidedly lower among barristers than in the general population; we have thus found a valuable result, though it may be worth while to subject it to a closer investigation. But ordinarily one might reasonably assume that the mortality in all age groups is lower, as is actually the case here.

Digitized by Google

English statisticians often use a modification of the method just described of calculating expected deaths; viz., the method of "standards" (in fact, the method of expected deaths can quite as well claim the name of a "standard" method). The difference between the two forms will be easily understood by using the same example as above. First, let us calculate the number of persons among whom, according to the general table of mortality and with the age distribution of the general population, 100 deaths would occur. Let us further calculate the number of deaths which would take place in this "standard" population, if the mortality were the same as observed among barristers and solicitors. The figures are as follows:

Age Period.	Standard Population.	Mortality per 1,000.		Expected Deaths.	
		General Population.	Barristers and Solicitors.	General Population.	Barristers and Solicitors.
25–34	2,625	6.38	4.72	16.75	12.39
35–44	2,040	10.94	7.21	22.32	14.71
45–54	1,475	18.67	13.63	27.54	20.10
55–64	959	34.80	27.89	33.37	26.75
25–64	7,099	14.08	10.53	100.00	73.95

According to mortality experience among barristers and solicitors, in a standard population approximately 74 would die instead of 100 according to the general rates of mortality. This is very nearly the same proportion as reached above. In the present case the two forms of comparison lead to nearly the same result, and this will generally be the case, if the age distribution in the special group is not much different from that of the general population. But on the whole the method described last is a little more complicated than the calculation of expected deaths, and in particular it is not applicable, if the age distribution of the deaths of the barristers and solicitors is unknown.

Still more complicated are other methods of procedure, such as the calculation of the mean duration of life or of the value of life annuities. In the present case we have the mortality experience for only a part of human life, and for ages above 65 we should be obliged to make estimates. Another way of comparison would be to calculate the number surviving at 65

Digitized by Google

out of a certain initial population. Such calculations may often throw light on the effect of the peculiar circumstances of the group under investigation; they ought never to be relied upon as the sole method of comparison, but may be used as supplementary to the simpler method of expected deaths.

X.

In economic statistics it will frequently prove of interest to have a common measure, a "standard," in order to get a clear view of the main result of the changes that may have occurred. Thus, when comparing the mercantile marine of two countries, or of the same country at two different periods, it may be useful to assign different weights to steamers and to sailing vessels, multiplying the tonnage of the steamers, for instance, by 3, thus taking into consideration the greater carrying power of steamers. Or in considering the livestock of a country, a cow being counted as 1, we may give a horse the weight of, say, $^3/_2$, a pig as $^1/_4$, a sheep as $^1/_6$, etc., and thus secure a rough estimate of the collective value of the livestock. A calculation of this kind may prove useful, but it cannot replace a close study of the original figures. The calculation may conceal changes which it would be of the highest interest to know thoroughly. Thus, in the history of agriculture, a decrease in the number of sheep might be offset by an increase of swine, and from such a computation we should learn nothing of this change. But having first studied the numbers in detail, we may get a comprehensive view by using some standard calculation. In any case the direct knowledge will always have a greater value than any abstraction, however ingeniously invented; and here, too, it may be laid down as a first principle always to keep the original numbers in sight as far as possible.

A most interesting example of a standard calculation is the index number, the familiar instrument of the economists. In trying to find the purchasing power of money at two different epochs we may calculate the budget of a certain model family, for instance that of a working man with a certain yearly income. If this budget at one time is $500 and the same quantities of all the commodities concerned at a later time would cost, for instance, $550, then the purchasing power

Digitized by Google

of money for this working man may be said to have decreased in the ratio $^{10}/_{11}$. But, if we used other budgets, we might get other results, so the comparison presents a defect, though probably the difference would generally be small. We are a little further from reality, if we calculate average values of the percentage variations of prices of different commodities, as is the case with most modern calculations of index numbers. In fact, we reduce the problem to that of using a budget in which each commodity enters with unity—or perhaps with a certain weight—a sort of abstraction from the actual budgets used above. It may easily be proved that a calculation of this kind cannot possibly give mathematically exact results. Let prices of two commodities at a given time be equal to unity, let the two prices at a later time be 0.80 and 1.25, and again at a third moment of time unity as before. Now taking our standpoint at the first moment, we find that the general index has increased at the second moment by $^1/_2$ (0.25—0.20) or $2^1/_2$ per cent. But from this moment till the third we have also an increase equal to $\frac{1}{2}$ (0.25—0.20) (the price of one commodity increasing from 0.80 to 1.00, or 25 per cent.). Thus in each of the two intervals prices have increased on an average $2^1/_2$ per cent. But the result of the combined movement from the first moment to the third will be nil, there has in fact been no change whatever. In practice this objection is not of great significance. Most economists would decidedly prefer the simple calculation of general index numbers by taking the arithmetic average of price changes instead of using a geometric average, such as recommended by the late W. Stanley Jevons, whose formula required operations with logarithms instead of with the numbers themselves. But at all events, whether we use one or the other formula, it will be necessary always to keep the various prices themselves in view. In the above example one commodity was becoming cheaper, the other, more expensive, and these facts will require close study before we try to find collective expressions for the changes.

Very often the characters we want to compare are not quantitatively measurable. If we investigate eye color, we may arrange the colors in a scale from blue to brown, and give the first rubric of the scale the weight one, the next two, etc.,

Digitized by Google

and calculate in this way the "average eye color" of particular groups. Thus, taking an instance from Galton's Family Records, we get the following results with regard to fathers and sons:

	Fathers Grouped According to Eye Color.	Average Number for Sons.
1	36	2.7
2	322	3.1
3	264	3.4
4	180	4.0
5	5	4.2
6	64	4.8
7	101	5.0
8	28	5.6
Total	1,000	3.7

It will be seen from this "contingency" table that the eye color of the sons corresponds to that of the fathers, in that the eye color of the sons shows a constant progression from blue to brown in the same order as that of the fathers, though there is a certain "regression" observable, the dispersion of the sons being much smaller than that of the fathers. Of course there is something unreal in such a comparison, the characters not being quantitatively measurable, and it would be better either to consider only the original data or to modify the calculation so as to conform more clearly to the observations. Professor Pearson proposes a distribution according to the exponential law, assigning to each group a mean distance from the average corresponding to this law. By this method the following distribution is obtained:

Group.	Fathers.	Sons.
1	1.0	3.8
2	3.3	4.3
3	4.8	4.8
4	5.8	5.2
5	6.3	5.6
6	6.5	5.7
7	7.0	5.8
8	8.0	6.4

It will be seen that the numbers on the whole have the same character as before; here, too, there is a smaller disp sion among the sons than might have been expected.

Digitized by Google

Such a method may be useful in dealing with certain problems, as, for instance, the results of examinations with peculiar marks for each discipline. At all events a detailed study of the reports themselves will always be indispensable as a check upon these calculations; in fact, low marks in one study, combined with very high marks in another, will give a different impression from that of a uniform grade in all studies.

XI.

One of the methods of comparison which have of late become popular among statisticians with mathematical training is that of correlation based on Bravais's formula. A simple example is offered by the age distribution of brides and grooms. If on a piece of plotting paper we put a point for each couple, the age of the bride being the abscissa, that of the bridegroom the ordinate, then we shall find that the points are not spread uniformly over the paper. but that they are clustered near a certain line, forming a sort of milky way. We may then by the method of least squares find the equation of the straight line which may be said to give the best idea of this milky way. This equation being $y-kx=0$, we have $\Sigma(y-kx)^2=$ a minimum, which will require $k\Sigma x^2=\Sigma yx$. If we further introduce the expression $\Sigma x^2=ns_x^2$, n being the number of cases, further $\Sigma y^2=ns_y^2$ and $\Sigma xy=ns_{xy}=nrs_{xy}$, where r is a quantity defined by the equation, we find the equation of the line in question:

$$y=kx=r\frac{s_y}{s_x}x$$

and the sum of the squares of the differences is:

$$\Sigma(y-kx)^2=ns_y^2, \text{ so that we have}$$
$$\mu^2=s_y^2\ (1-r^2).$$

If $r=1$, all the points will lie on the straight line; the smaller r is, the greater will be the mean error; when there is no correlation, we shall have $r=0$. This quantity r is called the coefficient of correlation, whereas the quantity k is called the coefficient of regression.

Digitized by Google

The formula of correlation will prove useful in all cases where the points are grouped nearly around a straight line, as is the supposition, and this will very often hold good, as in the case supposed of the age distribution of brides and grooms. Still we must not forget that this formula removes us somewhat from the original data and that it does not relieve us from the necessity of making a close investigation of these observations. On the whole, the formula of correlation does not introduce any new principle; by tabulating and grouping the observations we can easily establish as a rule the fact of correlation without the use of the formula. To take an example from Yule's *Introduction to the Theory of Statistics* (1911), the percentage of population in receipt of poor law relief in 38 English Poor Law Unions of an agricultural type is correlated with the average weekly earnings of agricultural laborers, and we find as the coefficient of correlation, $r = -0.66$. But it is unnecessary to make this calculation. Grouping the districts according to percentage of poor law relief, we find the following numbers, which tell us, without any long computation, how relief and wages are related:

Number of Districts.	Per Cent Poor Law Relief.	Average Wages (Shillings).
9	1.90	17.88
10	3.21	16.21
10	4.30	15.00
9	5.27	14.75
38	3.67	15.95

These numbers give us a perfectly clear idea of the connection between wages and poor law relief. The coefficient of correlation will tell us nothing which cannot be seen from an inspection of the original numbers. But in the illustration the amount of poor relief is influenced not solely by the average wages, but also by other influences, for in each group there are conspicuous deviations, and to explain them other causes must be found. The coefficient of correlation teaches us this and no more.

To take another example from Mr. Yule's book, what is the correlation between the fertility of mother and the fertility of

Digitized by Google

daughters in the British peerage? In each marriage of the mother generation one married daughter was chosen, and the number of her children put down, only marriages with a duration of at least 15 years being taken. It will be seen that the numbers are not quite homogeneous, in the first generation each marriage taken into consideration being fertile and with at least one daughter, in the second generation all marriages being counted.

We have the following numbers:

Mothers.	Daughters.			
Number of Children.	Average Number of Children.	Actual Number.	Expected Number.	Actual Number per 100 Expected Children.
1.....	3.2	167	230	73
2.....	3.5	201	247	81
3–4.....	3.9	905	1,006	90
5–6.....	4.1	1,087	1,145	95
7–8.....	4.8	976	889	110
9–10.....	5.1	655	555	118
Above 10.....	5.6	344	263	131
Total.....		4,335	4,335	

It will easily be seen how these calculations have been made; they are in fact quite elementary; the mean error in each group can be found without difficulty. In this method of calculation we have the advantage of being in close contact with the original observations.

Most of the problems which we meet in statistics of families are of the same elementary nature, and we shall gain very little in dealing with them by the more complicated processes. We shall always have to divide the observations in order to find the frequency in question in each group, and, if these groups are too small, we may combine the results by the method of expected cases or by some other simple method of calculation. The dependence of mortality of children on the age of the mother or the number of the birth, etc., or the rate of mortality in "phthisical" families is an example. The main difficulty is perhaps to find the moment in which the persons concerned enter into observation, for instance, when a death from phthisis occurs in the family. The following example is

Digitized by Google

taken from Lundborg, *Medizinisch-biologische Familienforschungen* (1913), giving details with regard to a certain epileptical disease (myoklonus) found in nine families with altogether 74 children. In this case we can take the family under observation after the birth of the first child suffering from the disease. The method will be seen from the table, in which 20 children who died under the age of ten have been left out of consideration.

Number of Family.	Number of Children in Family.	Number Diseased.	Distribution of Children (the Affected Ones Signified by x).								
			1	2	3	4	5	6	7	8	9
I	6	3	x	..	..	x	x	..			
II	8	1	..	x	..	..	..	..	..	..	
III	6	2	..	x	..	..	..	x			
IV	9	3	..	..	..	x	x	..	..	x	..
V	9	1	..	..	..	..	..	..	x	..	..
VI	5	2	..	x	..	..	x				
VII	6	2	..	..	x	..	..	x			
VIII	4	2	x	..	..	x					
IX	1	1	x								
	54	17									

In the first family the oldest child suffered from the disease; there were five children under observation, two of them being affected; in the next family the second child was diseased, but of the six children born subsequently none was affected, etc. We obtain the following result:

Number of Family.	Number of Children.	Cases of Myoklonus.
I	5	2
II	6	0
III	4	1
IV	5	2
V	2	0
VI	3	1
VII	3	1
VIII	3	1
IX	0	0
	31	8

Thus about one out of four children suffered from this disease. Weinberg proposes another scheme, which involves a somewhat more complicated calculation, leading, however, to substantially the same result.

Digitized by Google

XII.

Sometimes the problem is complicated by the absence of certain essential data; it is then frequently impossible to draw a safe conclusion. The question of the "handicapping" of first born children having been discussed of late in an interesting article in this Journal* by Louis I. Dublin and Harry Langman, I shall choose that problem as an example. On account of its great importance I shall deal with it at some length, discussing some Danish statistics of persons suffering from tuberculosis.

Number in Family.	Total Children.	Number Affected with Tuberculosis.	Per Cent. Affected.
1	3,522	988	28
2	3,344	713	21
3	3,043	568	18
4	2,619	427	16
5	2,161	271	13
6	1,731	198	11
7	1,296	113	9
8	962	81	8
9	636	46	7
10	445	48	11
11 and over	841	69	8
Total	20,600	3,522	17

Readers of this Journal will easily understand how this table was constructed, by consulting the paper I have cited. Seven hundred and thirteen second born children, for instance, were affected, their first born sisters and brothers will be found among the 3,522 first born, this number also including 988 first born affected with tuberculosis, 568 first born brothers or sisters of third born persons affected, etc.

There is a strong temptation to conclude that the first born, with their 28 per cent. affected have a much greater risk of being attacked by tuberculosis than the second or third, not to mention the later children. But it is evident that these observations are one sided, and that essential details which ought to have been included are omitted from the tabulation. If we inquire what is the probability of children with no sisters or brothers being affected, we find that it is one, all the children

*"On the Handicapping of the First Born" in American Statistical Association, *Quarterly Publications*, XIV (1915), 727-735.

Digitized by Google

observed in this group being in fact affected. The case resembles very much the problem of constructing a mortality table from observations of deaths only. The present statistics fortunately enough can be separated according to the number of the birth and the size of family, so that we can form the following table:

Number of Affected Child in Family.	Number of Children in Family of the Affected.											
	1	2	3	4	5	6	7	8	9	10	Over 10	Total
1	178	155	170	128	107	74	52	45	26	19	34	988
2	..	146	134	106	100	77	44	45	22	12	27	713
3	..	..	120	109	85	68	56	51	32	19	28	568
4	..	..	..	115	63	81	60	33	31	18	26	427
5	..	..	..	..	75	54	35	38	19	23	27	271
6	..	..	..	..	..	81	43	29	11	12	22	198
7	..	..	..	..	..	..	44	36	13	7	13	113
8	..	..	..	..	..	..	..	49	12	5	15	81
9	..	..	..	..	..	..	..	..	25	6	15	46
10	..	..	..	..	..	..	..	..	..	28	20	48
Over 10	..	..	..	..	..	..	..	..	..	..	69	69
Total	178	301	424	458	430	435	334	326	191	149	296	3,522

It will be seen from this table that 178 affected belonging to families with only one child have to be left entirely out of consideration, there being no other group of observations comparable to them. Consequently we have 810 first born comparable to 713 second born. Likewise we can compare 568 third born to 713−146=567 second born, etc. We can therefore arrange the observations as follows:

Number in Family.	Number Affected.	Number in Family.	Number Affected.	Expected Numbers.
1	810	2	713	712.1
2	567	3	568	561.6
3	448	4	427	420.3
4	312	5	271	305.8
5	196	6	198	219.8
6	117	7	113	147.3
7	69	8	81	99.6
8	32	9	46	58.8
9	21	10	48	37.6
10 and over	60	11 and over	69	..
Total	2,632		2,534	—

The expected numbers have been calculated by distributing 301 affected persons belonging to families with two children

Digitized by Google

uniformly on each number, further a third part of 424 persons in families with three children on each of the three birth numbers concerned, and so on, for instance

$$\tfrac{1}{2} \cdot 301 + \tfrac{1}{3} \cdot 424 + \tfrac{1}{4} \cdot 458 + \ldots = 712.1$$

The two columns in the above table would be homogeneous, if there were no other defects (as we shall see directly that there are). On examination the numbers of affected are found to correspond with the expected numbers fairly closely; there is no considerable difference with the single exception of the first born, who indeed seem a little "handicapped" compared to the later children. But the numbers are not quite homogeneous. To show this, let us suppose that the year of observation is 1910 and that each affected person was 30 years old, so that all the persons affected with tuberculosis in the table were born about 1880. Now the marriages of the parents of these persons are from varying epochs. Probably the parents of the first born generally married in 1879, of the sixth born perhaps about 1869. Comparing now in the table on p. 48 the numbers of the first and of the sixth child in families with at least six children, we find 250 against 198. But these 198 belong to marriages which are, say, 10 years older than those to which the 250 first born belong. But since the population has been increasing steadily, we cannot compare the numbers till we have made an allowance for this increase. Thus the difference between the two numbers will be much reduced. It seems probable, however, that we should find some little difference in favor of the later born children; it must be remembered that the first born children generally have a comparatively high infant mortality, so that proportionally few will survive,—a fact the correction for which tends to increase the difference. At all events the "handicapping" of the first born, as far as tuberculosis is concerned, must be insignificant compared to the enormous differences indicated in the table (p. 47).

This example shows how very difficult it is to handle one-sided observations. We are safe only when we have data not only on the numbers affected by tuberculosis but also on the whole population among which these cases of disease have been observed. What is wanted is a fraction, the numerator of

Digitized by Google

which gives the number of cases of disease, deaths, etc., and the denominator, the total number exposed to risk. Thus only can we draw a safe conclusion. If these required elements are not available, we must find whether and how far the defects in the statistics will invalidate our conclusions, by taking the various defects into consideration, as described above (pp. 17, ff.). This process, as we have seen in the present case, is often a very difficult task.

XIII.

Returning now to the problem of dealing with observations of characters which are not quantitatively measurable, I shall borrow some figures from a paper by Professor Pearson in *Biometrika* (III, 1904). One thousand seventy six pairs of sisters in a school were characterized by the teacher according as they were "quick," "good natured" or "sullen." The figures were as follows:

99 pairs, both sisters quick.
498 pairs, both sisters good natured.
60 pairs, both sisters sullen.
177 pairs, one sister quick, one good natured.
77 pairs, one sister quick, one sullen.
165 pairs, one sister good natured, one sullen.

Here again it seems to me that the data require only an elementary treatment in order to give full evidence as to the influences concerned. We have among 2,152 children 452 reported as quick, 1,338 as good natured and 362 as sullen (or 210, 622, 168 per 1,000): consequently according to the calculus of probabilities 44 per 1,000 of the pairs should be both of them quick, 261 should have one quick and one good natured sister, etc. We can construct the following table:

	Expected Number.	Actual Number.
Both sisters quick	48	99
Both sisters good natured	416	498
Both sisters sullen	30	60
One sister quick, one good natured	281	177
One sister quick, one sullen	76	77
One sister good natured, one sullen	225	165
Total	1,076	1,076

Digitized by Google

We see at once that the sisters are very often alike, 657 pairs having the same temperament against 494 as expected. It is not difficult to find the mean error in order to judge how much of the difference between the expected and actual cases is due to other causes than chance variation.

Professor Pearson prefers another way of dealing with this problem, arranging the experience in the following table:

Second Sister.	First Sister.			
	Quick.	Good Natured.	Sullen.	Total.
Quick	198	177	77	452
Good natured	177	996	165	1,338
Sullen	77	165	120	362
Total	452	1,338	362	2,152

We find thus each pair twice, in 77 cases one sister is quick, the other sullen; these are tabulated as 77 quick sullen and 77 sullen quick, etc. Giving now each group a separate weight we may calculate a coefficient of correlation. The result will be as above, but the calculations will require much more time than the more elementary procedure, and the results do not present any clearer view of the influences in question. We may simplify the table by reducing the number of columns, for instance, by adding the good natured and sullen, so that we get the following result:

Second Sister.	First Sister.		
	Quick.	Good Natured or Sullen.	Total.
Quick	198	254	452
Good natured or sullen	254	1,446	1,700
Total	452	1,700	2,152

But even thus we shall have more calculations than in the above elementary solution of the problem.

Sometimes even these contingency calculations increase the distance between the original observations and the calculated results to such an extent that we should decidedly

Digitized by Google

prefer the elementary calculations. Thus, in a recent review in the *Journal of the Royal Statistical Society* (Jan. 1916) of *The Material Culture and Social Institutions of the Simpler Peoples*, by Hobhouse, Wheeler, and Ginsberg, it is objected that a coefficient of contingency ought to have been calculated instead of confining the description of the conditions of these people to the simple statement that 47 per cent. of the "lower Hunters," 25 per cent. of the "higher Hunters," and 10 per cent. of the lower "pastoral" group fall into the category, "government slight or nil." It seems to me that this comparison is quite sufficient and that it would be waste of time to proceed to further calculations.

If I am right in these remarks, statistics of this kind can generally be treated in quite an elementary way, and no new results can be found by using more complicated methods. The principal problem will always be to find the relative frequency of the event or the character which we want to investigate, and first of all our object must be not so much to prepare refined statistical methods as to provide useful observations. The special nature of the observations may then lead to modifications of the formulas, to a more complete and refined theory, but at present we are more in need of statistical data than of theoretical investigations. This may sound curious, in view of the immense mass of details which are published every day by the numerous statistical institutions all over the world, but to a great extent all these reports are repetitions, so to speak, of older investigations, most of them made in a single mold. We want statistical observations covering new fields, and here, as shown above, an enormous amount of work remains to be done. Of course we cannot do without these myriads of statistical volumes; they have, at least the great bulk of them, their local claim to exist;—but beyond these reports numerous problems are waiting to be solved and the solution will require much patience and much careful work in gathering the necessary materials. As far as I can see, the future of statistics will depend on the energy with which these problems are taken up.

Digitized by Google

COMMENT ON WESTERGAARD'S "SCOPE AND METHOD OF STATISTICS."

By Warren M. Persons, *Colorado College.*

The reading of Westergaard's article on "Scope and Method of Statistics" makes me hope for the early translation of the manuscript and publication in English of his work on statistical theory. Particularly, I would like to have access to Westergaard's treatment of the probable error in time and geographic series, his discussion of the Pearsonian coefficient of correlation and his use of elementary methods of analysing data.

The present article leaves one doubtful concerning several points. For instance, at the end of Part I Westergaard says, "Consequently, there will be in the future an *embarras de richesses;* the difficulty will be not so much in gathering material as in mastering it and perusing all these masses of reports which have been stored in the archives and on the book-shelves of the statistical offices." In the concluding paragraph of the article he says, "First of all our object must be not so much to prepare refined statistical methods as to provide useful observations . . . at present we are more in need of statistical data than of theoretical investigations." In fact, we need both. If we have at hand the calculus which will enable us to derive the meaning from data of a certain kind, such a calculus will, first, encourage us to secure the necessary data and, second, determine the form of the data to be secured.

Westergaard does not correctly describe the economic statistics available today when he refers to them as "scanty material" of "doubtful reliability." Statistics of interest rates, exchange rates, production of commodities, prices, railroad earnings, clearings and the like are neither scanty nor unreliable. It is true we need more data, well classified; we also need a generally accepted calculus for treating them.

With Westergaard's characterization of the Pearsonian coefficient of correlation I emphatically disagree: "The coefficient of correlation will tell us nothing which cannot be

Digitized by Google

seen from an inspection of the original numbers." Recently I have been working on a problem involving the following question: Is the maximum correlation between steel production and prices and clearings obtained when concurrent items are paired or for a lag in one of the series? Such a question can be answered only by the application of a device such as the coefficient of correlation which enables us to measure the degree of correlation.

Representative investigations should, as Westergaard says, be used more frequently. Methods of interpolation and the theory of error must of necessity be developed for the treatment of the data thus secured.

My experience accords with Westergaard's that "it is possible to apply the binomial law of frequency to many fields of vital statistics," that in economic statistics the law is more difficult to apply, but that "the typical frequency curve in all vital, social, or economic statistics is always the binomial one." I agree that search for a formula which fits accurately a statistical series is usually unnecessary and that the securing of such a function is of use mainly, if not exclusively, for interpolating values. It is obvious that by using a sufficiently large number of constants a function may be found to "fit" with a high degree of accuracy any statistical series.

Westergaard's caution that we should "keep the original data in view" is one that every investigator should hold in mind.

The publication of Westergaard's book in English will be warmly welcomed by American statisticians.

Digitized by Google

COMMENT ON WESTERGAARD'S "SCOPE AND METHOD OF STATISTICS."

By Julius H. Parmelee, *Washington, D. C.*

Westergaard's treatise is an inspiration to statisticians; first, because it breathes a wholesome optimism as to the mission of statistics in the settlement of social problems; secondly, because it offers definite suggestions for the utilization of mathematical laws in statistical work.

The thirteen sections of the treatise have a marked German flavor, in their painstaking thoroughness, their attention to a wide mass of detail, and their fidelity to what may be termed a scientifically statistical atmosphere. Westergaard descends to no defense of statistical method or of the science of statistics as a whole. He expends little energy in the time-honored discussion whether statistics is a science or a method. He accepts things as they are, and having cleared the way by a characteristically painstaking historical discussion, plunges boldly into his thesis, which is, in brief, that while certain mathematical laws must be recognized and utilized, yet they should be applied with due regard to actual conditions of social development.

This, it seems to me, is Westergaard's particular contribution in the present work, even more than his clarion call to statisticians to be up and doing, or his interesting discussions of the many and varied social phenomena of which he shows so remarkable a grasp. Let me illustrate from various sections of the article.

The several laws described by Westergaard are shown each to have a place in statistical method, dovetailing with other laws in such manner as to enable statisticians to arrive at definite results and conclusions. Take the law of error, which is both based on, and limited by, the calculus of probabilities. By use of these mathematical factors, which were largely unknown to, or at least unused by, earlier statisticians, the present-day student of social movements can often test the accuracy of his facts almost as closely as can the chemist his

Digitized by Google

laboratory reactions or the biologist his developments in the embryo. Granted a sufficiently large group of observations, and a sufficiently cautious regard for all the complicated social factors involved, the statistician can plot the progress of society with what practically amounts to mathematical nicety. If the earlier statisticians failed to make adequate use of these laws, it was because they did not understand their full significance. Further, as Westergaard points out, where they did utilize the laws they failed to appreciate that their application to social phenomena required an understanding of all the conditions involved, and of the variability of such conditions. It follows that the modern statistician can appeal to a wider range of collateral sciences for assistance in adjusting his laws, such as physiology, biology, psychology, and the like.

Again, the law of interpolation frequently enables the student of statistics to fill out gaps in his data almost to the final fraction. In common with other laws, Westergaard points out that this law should not be regarded as something working wholly apart from human developments, but, as he says, "this interpolation should keep as near real life as possible, *taking everything into consideration* which might prove of importance." The italics are mine, and serve to emphasize the author's insistence, more often implied than stated, upon the absolute dependence of the statistical student on things as they are. He should not be a scholar put into his laboratory to grind out laws and methods, but a live student of human nature and human customs, adjusting his statistical theories in accordance with the *mores* of his time.

Closely related to the calculus of probabilities, in fact, almost a part of it, is the law of expectation, or actuarial law. This vital law of society enables the statistician to judge whether society is progressing or retrogressing, whether the death rate is falling and the birth rate rising, or vice versa, whether the laws of heritage are or are not operating in accordance with theory and schedule—in fact, rightly applied, it gives the social student a marvelous motion picture of society in its dynamic phases. This indicates the importance of the law, and the absolute necessity of its being accurately calculated and applied to actual conditions.

Digitized by Google

In fine, Westergaard holds that the earlier statisticians were really the predecessors of modern social students, even when they devoted themselves exclusively to the mathematics of probabilities, or to the intricacies of political arithmetic, or more particularly to the description of the comparative conditions of different states and countries. Even the last group, who were ridiculed as "Tabellenknechte," or slaves of the table, were in their feeble way working toward the same goal as the modern statistician, namely, toward an accurate picture of the condition of society. So, today, the statistician should not be a mere slave to his statistical columns, with nose buried between the covers of his logarithmic tables, but should be a wide awake and sympathetic social alchemist, applying ascertained facts to the social problems of his times. Yet Westergaard would have the statistician appreciate his limitations. As a social student he may apply the facts he has ascertained as a statistician, but the statistician's primary duty is to secure the facts. He "must be content if he can show that certain groups of numbers show marked differences, leaving it" to others to explain the differences. This does not inhibit him from utilizing, as a social scientist, the facts he has gained as a statistician.

It must be clear, from this brief analysis of Westergaard's thesis, that he has made society his lifelong study and devotion, for his volume breathes the joy and zest underlying the study of social progress by means of accurate scientific methods, and the book in unabridged form can hardly fail to be an inspiration both to statisticians and students of social development the world over. If there are criticisms to be offered of his article, the chief is his utilization of certain theoretical studies by biologists and others that would seem only remotely akin to the working of definite laws; but such defects as these seem to me fully offset by the sound optimism of his principal thesis and conclusions.

Digitized by Google

COMMENT ON WESTERGAARD'S "SCOPE AND METHOD OF STATISTICS."

By Carl J. West, *Ohio State University.*

It is often desired to estimate the value of a population for intercensal years. How is it to be done? Since a direct enumeration is out of the question some indirect method must be employed. As a basis for an estimate it may be assumed, for instance, that the ten year increase has proceeded by equal annual additions, or that there has been a constant annual rate of increase. Before the question of the method to be used can be definitely answered, the statistician must be familiar with the advantages and the disadvantages of each method.

A second problem of importance to the statistician making use of census data is that of distributing the excessive numbers appearing at the even ages and especially at certain ages, at age 40 for instance. No method of reënumeration can assist in the solution of this problem because the error is essentially inherent in the only practicable methods of enumeration. It might among other possible assumptions be assumed that the excess at age 40 was drawn from ages 41 and 42, or that a graphic solution is the best. Ordinarily not until quite complicated methods are made use of in removing these and other irregularities can the census figures be said to show the approximate age distribution.

It is not necessary to bring forward repeated illustrations to show that a statistician must be more than a mere enumerator, that even in the most simple statistical work some theory is absolutely necessary. It must be equally clear that instead of much of the personal and tentative methods of present day statistics there should be a consistent body of statistical theory which would be generally understood and applied. What is needed is more articles and more books and treatises presenting the theory and methods of statistics in a simple style. And further and above everything else the *theory and methods*

Digitized by Google

should be presented as a distinct science—the science of statistics. Only when statistics is generally recognized as a science worthy of equal rank with physics and chemistry and biology will it secure the proper respect and recognition without which the labor of statisticians is partially in vain.

Let us examine the article, "Scope and Method of Statistics," with the purpose of discussing Westergaard's idea of the necessity for the development of an exact science of statistics.

We begin with the theory of representative statistics. As a class economists have hesitated to make use of representative statistics, perhaps because little use can be made of such material without reference to certain very practical points of method. When can a selection be considered as representative, as typical of the whole class of individuals from which it is drawn? How is the significance of a variation in a sample or selection to be determined? are some of the questions which confront the statistician. "What remains to be done is simply to develop a theory of these representative enumerations, stating the limits of the deviations from the true proportions and showing how to approach as nearly as possible to the truth. In fact in many cases it will be practically impossible to do without representative statistics."

The matter of inaccuracies in the data is always one of great concern to the careful statistician. Note the statements of Dr. Westergaard: "a certain pessimism is often encountered among official statisticians and economists who attempt to draw conclusions from statistics. It is maintained that the inaccuracies are often so great that it is impossible to get any reliable results. Here, then, is another important problem in the theory of statistics, viz., to determine the significance of the inaccuracies, to state to what extent it is possible to draw conclusions from statistical date in spite of their imperfections. Here it would be most useful to form a *theory of the applicability of imperfect data.** For it can easily be shown that even extremely incorrect data may under certain circumstances allow of perfectly safe conclusions."

The author's discussion of the "law of error" in Section IV

*Italics are mine.

Digitized by Google

is of far reaching significance. To illustrate the nature of the discussion suppose of 400 patients with a certain disease 200 have been treated with a serum and 200 have not been so treated. Of the first class suppose further that 125 recover and of the second class only 100 patients recover. Is the observed excess of recoveries among the patients of the first class sufficient to show the efficacy of the serum treatment? In a certain community the ratio of male to female births in one year is 110 to 100 instead of the more usual ratio of 105 or so to the 100. Can one, because of this larger ratio, be certain that some cause is present which tends to increase the relative proportion of male births or is the observed excess merely a chance fluctuation, just as when tossing coins a run of heads may occur? As the basis for a fundamental theory of these variations Professor Westergaard advocates the use of the standard deviation, or "mean error," of a binomial series taken in connection with a normal law, or "law of error" as he terms it. He presents a strong case for the use of this method of measuring variations. It is of course true that not all data even approximate to the binomial law, but actual trial and experimentation show that even so the standard deviation constitutes a fairly reliable measure of variability. Altogether there appears to be a no more adaptable basis for a general theory of variability than that furnished by the standard deviation.

The idea of the author that statistical data can be broken up into a series of normal distributions is of extremely high importance as a theoretical basis for the study of a statistical distribution. The analysis of a given set of statistical measurements by identifying the component variations and values as constituent, normal distributions is one of the most significant ways imaginable for extracting information from stubborn data. It gives an arithmetic background and general basis for the work. The statistician with the theory of the normal distribution in mind feels more confident, feels that he has a standardized viewpoint, so to speak. Probably no statistical distribution can be said to have yielded all its information until it is broken up into the constituent "laws of error," or normal distributions. It may here be of interest

Digitized by Google

to define, in statistical terminology, a normal distribution. *A series of measurements is said to obey the normal law when each measurement, X, is equal to the sum of a relatively large number of smaller elements where each constituent element has a value chosen at random.** Thus the height of a person is the sum of the lengths of certain bones, of the widths of certain cartilages, and each of these elemental values is, within rather narrow limits, about as likely to take one value as another. Hence it would be anticipated that a distribution of height would be normal. The just mentioned value to statistics of basing the theory on a normal distribution is not lessened by the fact that in practice it is ordinarily not easy to split a distribution into its constituent normal distributions.

The author shows the true instinct of the statistician when he says at one place "On the whole we may here lay down the principle that this interpolation *first of all should keep as near real life as is possible, taking everything into consideration which might be of importance";*† at another place, "The best means of treating the changes of a variable in a series of observations is the method used by Daniell Bernoulli and Duvillard. They used the fiction that all variations were continuous so that it was possible to use the differential calculus. . . . This method is of great value as it *simplifies many problems extremely.*† . . . The main difficulty at present seems rather to consist in providing sufficiently complete and correct observations than in dealing with these materials afterwards."

The tendency of the author to systematize, to reduce to general principles, is shown by the sentence, "Still more complicated are the problems of economic statistics for here we frequently have *one dimension more,*† we have to deal not only with numbers of quantities but with their value in money as well."

Again, "Here we have many methods of procedure. *In determining which method is the best it is a good principle always to keep as near as possible to the original observations, never leaving them out of sight, with as few intermediate links as possible."*

* Edgeworth, "Law of Error," Cambridge Phil. Trans., 1905.
† Italics are mine.

Digitized by Google

On the formula of correlation we find, "The formula of correlation* will prove useful in all cases where the points are grouped nearly around a straight line; . . . *still we must not forget that this formula removes us somewhat from the original data and that it does not relieve us from the necessity of making a close investigation of these observations.* On the whole the formula of correlation does not introduce any new principle; by tabulating and grouping the observations we can easily establish as a rule the fact of correlation without the use of the formula." The author does not mention the value of a diagram—a scatter or correlation diagram, and he is apparently not familiar with the properties of the correlation ratio, but his viewpoint is fundamentally that of seeking the underlying principles of the science and of discussing each proposed method with reference to those principles.

Again with the same purposes in mind he writes with reference, apparently, to Professor Pearson's elaborate methods for determining correlation for non-measurable characteristics such as hair color, disposition, material wealth, and so on, "but at present we are more in need of statistical data than of theoretical investigations. This may sound curious in view of the immense mass of details which are published every day by the numerous statistical institutions all over the world, but to a great extent all these reports are repetitions, so to speak, of older investigations, most of them made in a single mold. . . . Of course we cannot do without these myriads of statistical volumes, they have, at least the greatest bulk of them, their local claim to exist;—but beyond these reports numerous problems are waiting to be solved and it will require much patience and much careful work in gathering the necessary materials."

It would perhaps be better to say that there is equal need of statistical data and theoretical investigations. For until the methods are clearly in mind the search for data is likely to continue to be a more or less haphazard groping in the dark. There must be some model, some plan, for the collection of data and there is nothing which can give a greater

*The coefficient of correlation alone is referred to here. C. J. W.

Digitized by Google

unity to the gathering of the material than can generally accepted and applied methods.

In conclusion it may be said that Dr. Westergaard in this article shows a thorough appreciation of the needs and difficulties of practical statistical work together with that scientific sympathy and spirit which makes all his work count towards the development of statistics into a distinct methodological science.

Digitized by Google

COMMENT ON WESTERGAARD'S "SCOPE AND METHOD OF STATISTICS."

By G. P. Watkins, New York, N. Y.

The special viewpoint of Westergaard would by many be called mathematical, or he would be more definitely described as one who thinks of statistics as an application of the mathematical theory of probabilities. The description is substantially correct. Its appropriateness appears in the article under discussion and would be borne out by an examination of the author's *Grundzüge der Theorie der Statistik.* But such a summary statement is seldom adequate. Some statisticians are statisticians, and some are mathematicians. Of course the implied opposition needs qualification. Indeed the most that need be claimed for it is some relative significance. The point of present interest is that Westergaard belongs with the statisticians. He brings to the aid of statistics certain mathematical conceptions, but he has no preference for the processes and expressions of algebraic mathematics and uses them only to serve his statistical purposes. This is not always true of the mathematical statistician. It is a very superficial view that supposes, because both mathematics and statistics are concerned with numbers, competence in the one involves competence in the other. Great familiarity with abstract numbers and great satisfaction in dealing with them do not promote skill in handling and interpreting numbers standing for quantities and quantitative relations of concrete things. That a statistician should be "quick at figures" and fond of equations and mathematical formulas is no more true than the other extreme view that only a physician can be competent in vital statistics, only a teacher in educational statistics, only a farmer in statistics of agriculture, etc. There is an element of truth in the second viewpoint. Concrete knowledge of the things to which statistical numbers relate is of great importance. The mistakes that are most commonly used to discredit statisticians (though they are not in fact very harmful) are due to lack of sufficient

Digitized by Google

knowledge of the concrete situations to which the numbers dealt with relate. But one can know something about farming without having spent years handling a pitchfork, and about the railroads without having been a brakeman, etc. Moreover, statistics has methods and a technique of its own —distinct from those of mathematics—with which professional familiarity is as necessary for the statistician as is the professional familiarity of the physician with disease, of the farmer with crops, etc.

The frequency and appropriateness of Westergaard's references to the necessity of keeping the original data in sight is a conspicuous characteristic of his article and the one that has suggested the foregoing remarks upon the fundamental concreteness of statistical numbers. A death rate, for example, should not be thought of as a mere ratio. Its meaning depends upon age, sex, race, etc. The "etc." is not a mere form. The statistician would like to differentiate and classify *ad infinitum* the deaths and the persons exposed to death, and find the correlation, if any, between the various pairs of classifications. The quality of statistical analysis is not secured by correctness of calculations, nor by appropriate application of mathematical theory and skillful manipulation of mathematical formulas. Accuracy of observations is more important. But statistical accuracy does not mean mathematical exactness. It means rather the reduction of biased error to a minimum, and also—but this may usually be taken for granted—the use of large enough aggregates so that unbiased errors cancel each other. The more or less selective character of observation, statistical and other, and the varied composition of concrete numbers—the two factors working together—constitute the great source of wrong statistical conclusions. The "probable error" in which the mathematical statistician is particularly interested is of comparatively little significance.

There seems to be no occasion to add anything to what Westergaard has to say of the significance of the probable error for statistical analysis. But there is one matter of terminology in relation to which I would express dissent. The reference to sample statistics as "representative statistics"

Digitized by Google

seems to me infelicitous. The adjective should refer to function and quality rather than to a situation and a numerical relation in which the quality is chiefly assumed, the only reason for alleging that samples are representative being that their selection is presumed to be unbiased. Moreover, we need the word in a truer sense for *representative numbers*, where the process by which they are obtained does have explicit reference to the function and quality in question. I refer to "averages" of all sorts, mathematical and unmathematical, weighted (or true) and unweighted (and apparently hap-hazard). Representative numbers, in the preferred sense, and relative numbers are the two great instruments of statistical analysis. Representative numbers are such as provide a condensed substitute, in some one or another use of the figures, for the complete aggregate. An average *represents* an aggregate. We try to obtain sample numbers that will represent the comprehensive aggregate where it is impossible to observe statistically the latter as a whole. But the statistical process in this case assumes rather than assures the representativeness desired.

One matter that is quite pertinent to the viewpoint of Westergaard's article seems to me not sufficiently developed. A statistical discussion of "probable error" should not stop at the application of mathematical theory. It should consider the *margin of error* in a broader sense. "Probable error" in the mathematical sense deals with accidental and unbiased and relatively small errors. But the statistician who has had to do with the interpretation of and compilation from enumeration and report schedules does not worry about such unrepresentativeness of his results as may be due to a large "probable" error. Most government statistical offices deal, not with samples, but with comprehensive or complete aggregates, which are also large enough to get the full benefit of the "law of large numbers." Where the "probable error" is negligible, however, the *margin of error* may be large. The practically important meaning of accuracy relates to the latter situation. The accuracy of a census may be tested by repeated sample enumerations, by known or determinable coloring or inefficiency in administration, by internal evidences

Digitized by Google

of duplication or omission, by the known functional relation of true results to figures obtained by other methods, etc. But all such errors are presumably biased and they tend to be too large to be left to chance cancellation. The problem is not mathematical.

I am not able to say whether Westergaard's "method of expected cases" is his original contribution to statistics or not. Perhaps the method is not in principle different enough from the "standard calculation," generally known in its application to the "correction of the death rate," to make the question of originality worth particular consideration. The "correction" method is substantially the inverse of the method of expected cases. At any rate the development and use of the latter by Westergaard affords an example that ought to influence other fields besides that of vital statistics. I obtained my conception of the importance of the method from a reading of his *Mortalität und Morbilität* (2d edition, 1901). Section 18 of chapter I introduces the idea, and the frequent apt use of the method throughout is not the least reason why the book deservedly takes the highest rank among statistical monographs. The method deserves a distinct and important place in general statistical theory. It is both simple and of broad applicability. I have found what is in effect the same method useful in studying street railway traffic increases. It should be of interest for the study of rates of growth in general. In allowing for the effect of annexed data upon per cent. increases the principle under whatever name must be taken into consideration. It should be noted that the method is not only elementary, as regards the mathematics involved, but that the required direction of attention is to the concrete.

Digitized by Google

THE FEDERAL FARM LOAN ACT.

By Walter B. Palmer, *Special Agent, Bureau of Foreign and Domestic Commerce, United States Department of Commerce.*

The Federal Farm Loan Act. The passage of the Federal Reserve act, which provided a more elastic currency system for the commercial needs of the country, has been followed within three years by the enactment of the Federal Farm Loan bill, which furnishes facilities for farmers to obtain credit on far more advantageous terms than have been offered to them heretofore. The latter bill is intended to place agriculture on an equal footing financially with other industries. It passed both branches of Congress by practically a unanimous vote. Considered from financial and economic standpoints the Federal Farm Loan act is one of the most important measures ever enacted by Congress.

Until the passage of the Federal Reserve act, in December, 1913, national banks were prohibited from making loans on real estate security, and farmers found it difficult to secure credit accommodations at reasonable rates of interest. The only security that farmers can give for a loan is usually real estate, and frequently they are charged exorbitant rates of interest, in violation of usury laws. Recent investigation by committees of Congress and the Department of Agriculture showed that the average rates of interest on farm mortgages, including commission, varied from 5.3 per cent., in New Hampshire, to 9 and even 10 per cent. in certain Southern and Western States; in some instances they were as high as 25 per cent.

The Federal Reserve act permits member banks to make loans secured by mortgages on improved farm lands, for periods of not exceeding five years, to the extent of only 25 per cent. of their capital and surplus. This makes available for such loans less than $500,000,000, whereas existing farm mortgages amount to over $2,000,000,000. Federal Reserve banks are organized principally for the benefit of the commercial and manufacturing classes. They extend or con-

Digitized by Google

tract their loans in response to the rise and fall in the supply of deposits. They must be prepared to return deposits on demand, and it is important for them to invest these deposits in only short time loans.

Federal Reserve banks have, therefore, shown little disposition to make loans on farm mortgages. Some state banks and savings institutions make such loans and permit them to run five years, but usually insist on notes payable on demand, so that the bank funds will not be tied up in case of sudden calls for the return of deposits. The farmer borrower is, therefore, in constant dread that he will be compelled to repay his loan at a time when it will cause him inconvenience or perhaps loss. When loans on farm mortgages are renewed the farmer must pay additional fees and frequently new commissions. In some states the mortgagor must pay taxes on the mortgage as well as on the land. In other states the mortgagee must pay taxes on the mortgage, which, of course, are reflected in the rate of interest.

The census of 1910 showed that the rural population of the United States then was 49,000,000 and the urban population (population of incorporated places, including New England towns, having 2,500 inhabitants) 42,000,000. The farming class, constituting about half the total population, needs loans for the purchase of additional land, the erection of farm buildings, the construction of roads, drains, and fences, and the purchase of machinery and stock. If farmers can obtain loans for such improvements on reasonable terms, agricultural operations will be greatly extended, and lands made more productive, with the result that farm products will be more plentiful. This will have far reaching effects, for with increased production consumers would be benefited by lower prices. Country life will be made more attractive and the tendency of farmers' sons to seek employment in the cities will be checked.

Of all the leading industries agriculture has been more hampered in obtaining credit for operating expenses than any other. According to the census, the value of farm products in 1909 amounted to $9,000,000,000, and of manufactured products $21,000,000,000, or over twice as much. But the

Digitized by Google

amount invested in farm lands, buildings and equipment, $41,000,000,000, was over twice the capital employed in manufacturing, $18,000,000,000.

Many farmers compelled to borrow money to carry on farming operations would have been ruined by the high rates of interest and commission charges that they have been forced to pay had it not been for the great increase in the value of farm property during recent years. The Federal census reports that during the 10 years from 1900 to 1910 the average value of farm lands in the United States increased from $15.57 to $32.40 per acre, or 108 per cent.

The percentage of mortgaged farms increased during these 10 years from 31 to 33.6 per cent., though this was partly due to the breaking up of Western ranches and Southern plantations into smaller holdings. The percentages of all mortgaged farms for which reports concerning mortgages were obtained by census enumerators are shown in the following table, these percentages being based on the combined total for farms free from mortgage and farms mortgaged in the different sections of the United States:

PER CENT. OF MORTGAGED FARMS.

Section.	1910.	1900.	1890.
United States	33.6	31.0	28.2
New England	34.9	34.1	28.2
Middle Atlantic	38.3	40.3	37.0
East North Central	40.9	39.4	37.6
West North Central	46.1	44.3	48.0
South Atlantic	18.8	16.8	7.4
East South Central	22.7	17.0	4.5
West South Central	30.6	18.2	4.8
Mountain	20.8	14.4	14.1
Pacific	36.8	27.6	28.7
The North	41.9	40.9	40.3
The South	23.5	17.2	5.7
The West	28.6	21.7	23.1
East of the Mississippi River	30.9	29.7	24.8
West of the Mississippi River	38.0	33.4	34.7

In every geographic division except the Middle Atlantic the proportion of farms mortgaged was greater in 1910 than

Digitized by Google

in 1900 and in every division except the West North Central the proportion was greater in 1910 than in 1890.

There is an abundant supply of funds which, with perfect security, could be used for investment in long term mortgage loans at low rates of interest, such as the savings of people earning wages or salaries, the proceeds of life insurance in the hands of widows and other beneficiaries, funds belonging to estates, minors, and wards in chancery, funds of insurance companies and benevolent societies, endowments of colleges, hospitals and other institutions, trust funds, and assets handled by receivers in bankruptcy. The custodians of such funds, which aggregate an enormous sum, have not usually made loans to farmers direct, but largely through investment companies charging high rates of interest and commissions.

The Federal Farm Loan act puts it in the power of those interested and those who have funds to invest to extend to farmers the credit at reasonable rates which they need as much as it is needed by merchants and manufacturers. The investment may be made in farm loan bonds, and the proceeds from the bonds used for accommodating farmers, who in the aggregate produce a great part of the wealth of the country. The bonds issued under federal supervision, protected by every safeguard, legally acceptable as security for trust funds, and exempt from all taxation, will offer a most attractive form of investment, and investors as well as farmers will be gainers.

All loans made under the Federal Farm Loan act will be made on an amortization plan, by which the principal is repaid in annual or semi annual installments during a period of from 5 to 40 years, interest payments being made at the same time. Experience teaches that a borrower is much more likely to meet his loan obligations without default if he pays the principal in small regular installments at fixed periods, instead of meeting the whole obligation after the expiration of several years.

The Federal Farm Loan Board has supervision over federal land banks, which, in turn, have supervision over national farm loan associations. Loans are made by the banks through the associations to members of the associations. The banks are authorized to issue bonds, secured by the first mortgages

Digitized by Google

they hold, and to lend to farmers giving first mortgage security the proceeds from the sale of the bonds.

The rate of interest on the mortgages shall not be more than 1 per cent. over the rate of interest of the bonds. For example, if the current rate of interest on farm loan bonds is 4 per cent., the bank may charge not to exceed 5 per cent. to farmers who borrow. This margin of 1 per cent. on the amount of unpaid principal is available to the bank for its expenses and earnings, but it will probably be reduced in practice. The act provides that the interest on bonds issued by the banks shall not exceed 5 per cent., so that the rate of interest paid by mortgagors will in no instance exceed 6 per cent., and this may be reduced by the dividends of the bank.

The national farm loan associations, which recommend loans by the federal land banks, are coöperative organizations, and are managed with a minimum of expense. Members of such associations may be expected to see that the best available men in their respective communities are selected as directors, and to keep careful watch lest unsafe loans be made with resulting personal loss to all members.

In Europe much has been accomplished by legislation that encourages coöperation. There are societies to organize cooperative institutions and lecturers who go about to spread the coöperative gospel. There are coöperative insurance companies, coöperative buying and coöperative selling organizations, coöperative markets, creameries, and breeding establishments and the like. In most European countries coöperative land mortgage and rural credit systems have been in successful operation for many years. The Federal Farm Loan act is but one factor in the vast field for coöperation on a national basis that lies before the American farmer, but it will greatly stimulate the coöperative spirit and will lead to the formation of other coöperative enterprises on a large scale.

The fact that American agriculturists were not provided with suitable credit accommodations has long been recognized, but the subject did not receive general attention until 1912. The passage of the Federal Farm Loan bill is the result of

Digitized by Google

four years of effort and publicity. Mr. David Lubin, who had been the American delegate to the International Institute of Agriculture, was invited to make an address on the subject of a rural credit system to the Southern Commercial Congress at its annual convention held at Nashville in 1912. He accepted on condition that the representatives from the various states should assemble a week prior to the convention and devote that time to the study of the subject. Twenty seven states sent representatives to the conference.

Mr. Lubin addressed the conference, explaining rural credit systems in foreign countries, and afterwards devoted much attention to the proposition of establishing such a system in the United States. The convention decided that a commission, composed of two men from each state, should go to Europe to study the subject in countries where such systems were in operation. The matter was taken up with President Taft, the Department of State, the governors of the various states and with organizations of farmers throughout the country, and the plan received general endorsement. It was endorsed in the platforms of the three major political parties in their national conventions in 1912.

Later, the agricultural bill passed by Congress contained provisions for a Federal Commission of seven members to be appointed by the President, "to coöperate with the American Commission assembled under the auspices of the Southern Commercial Congress." The American Commission, composed of two men from each of 35 states and four Canadian provinces, and five members of the Federal Commission sailed from New York in April, 1913. At Rome they were received by the King and Queen of Italy at the home of the International Institute of Agriculture. Through delegates to the institute, plans were made for a tour of 14 European countries. The commissions were aided by every facility the various governments could offer, and they returned to the United States in July, 1913.

A great mass of valuable, authoritative data was collected by the two commissions. Elaborate reports were made by the commissions—by the Federal Commission to Congress and by the American Commission to the governors of all

Digitized by Google

states,—and the data and reports were published as Senate documents, and widely distributed, especially among farmers and agricultural organizations and journals. This publicity aroused wide interest in the proposition. The measure was strongly endorsed by President Wilson, as it had been by President Taft. President Wilson's annual message to Congress, in December, 1913, was largely devoted to the subject.

The report of the Federal Commission, containing the bill recommended, was presented, in January, 1914, to the Senate by Senator Duncan U. Fletcher, of Florida, and to the House by Representative Ralph W. Moss of Indiana. Senator Fletcher was chairman of both commissions that visited Europe, and Mr. Moss a member of the Federal Commission. Senator Fletcher was President of the Southern Commercial Congress in 1912, when the movement started, and has been since then.

The agricultural appropriation bill enacted in 1915 provided for a joint committee on rural credits, to consider all bills on the subject and to report to Congress. This joint committee consisted of Senators Owen of Oklahoma, Hollis of New Hampshire, Gore of Oklahoma, Smith of Georgia, Nelson of Minnesota, and Brady of Idaho; and Representatives Glass of Virginia, Phelan of Massachusetts, Lever of South Carolina, Moss of Indiana, Hayes of California, and Hawley of Oregon. Representative Carter Glass was made chairman of this joint committee, Representative Ralph W. Moss of a sub-committee on personal rural credits and Senator Henry F. Hollis of a sub-committee on land mortgage loans. Through hearings before these committees, and by means of a wide distribution of literature on the subject, the whole country became interested as rarely before in a measure, and a realizing sense of the need for legislation became nation wide.

The draft of a revised bill and a report by the sub-committee on land mortgage loans were adopted by the joint committee on rural credits and presented to Congress in January 1916. The report, signed by Senator Hollis for the sub-committee, concluded with the following summary:

It is believed that the system of land banks outlined in the proposed bill affords a safe and attractive farm loan bond

Digitized by Google

for the investing public; low interest rates, long term mortgages, and easy payments for the farmer; low cost of administration; simplicity of organization and operation; adaptability to the needs of every section; and stimulation to the spirit of generous coöperation among farmers.

The Senate and House referred the bill to their committees on banking and currency. The bill, when reported with a few changes, was passed in the Senate by a vote of 57 to 5 and in the House by a vote of 295 to 10. The conference committee consisted of Senators Owen, Hollis, Fletcher, Nelson, and Gronna, and Representatives Glass, Phelan, Moss, Hayes, and Platt. The conference report was agreed to by viva voce vote in the Senate, and with only 12 votes recorded against it in the House. The bill was approved by President Wilson on July 17, 1916.

The Federal Farm Loan Board. A bureau of the Department of the Treasury, called the Farm Loan Bureau, will execute the provisions of the act, under the general supervision of a Federal Farm Loan Board, with headquarters in Washington. This board consists of five members, of whom one is the Secretary of the Treasury, who is chairman *ex officio*. The other four members are appointed by the President of the United States, by and with the advice and consent of the Senate, and not more than two may be appointed from one political party. One of the four is designated by the President as the Farm Loan Commissioner, who is the active executive officer of the board.

One of the four members appointed by the President serves for two years, one for four years, one for six years, and one for eight years, and thereafter each of these four members will serve for eight years, unless sooner removed by the President for cause. These four members receive a salary of $10,000 annually, and must devote their entire time to the business of the board, and not engage in any other banking or mortgage loan business.

The Federal Farm Loan Board has power:

To organize and charter federal land banks, to charter national farm loan associations and joint stock land banks, and to exercise general supervisory authority over them.

Digitized by Google

To review and alter the rates of interest to be charged for loans made by federal land banks, these rates to be as uniform as practicable.

To make rules respecting the charges to borrowers for expenses in appraisal, determination of title and recording.

To grant or refuse to federal land banks or joint stock land banks authority to make any specific issue of farm loan bonds.

Upon request of the Federal Farm Loan Board, the Secretary of the Treasury is authorized to make deposits for the temporary use of any federal land bank, at current rates of interest for such deposits, to be secured by farm loan bonds or other collateral satisfactory to him. Certificates of such deposits held by him shall be redeemed at his discretion. The aggregate of Government funds thus deposited may not exceed $6,000,000 at any one time.

The sum of $100,000, or so much thereof as may be necessary, is appropriated for carrying into effect the purposes of the act. The salaries and expenses of the members and employees of the Federal Farm Loan Board, and of a farm loan registrar for each federal land bank district, and of land bank examiners are to be paid by the United States. No funds of the United States may be used for paying any other salaries or expenses of the federal farm loan system. The officers and employees mentioned are excepted from civil service rules, but the President has the power to place them in the classified service.

It is the duty of the Federal Farm Loan Board to prepare amortization tables which are to be used by national farm loan associations, federal land banks, and joint stock land banks, to prepare forms on which such associations and banks report to it quarterly, and annually to make a full report to the Speaker of the House of Representatives, to be printed for the information of Congress.

Federal Land Banks. The Federal Farm Loan Board is directed to divide the continental United States, excluding Alaska, into 12 districts, no district containing a fractional part of any state, and to establish in each district a federal land bank, with its principal office in such city as the board

Digitized by Google

shall designate. Subject to the approval of the Federal Farm Loan Board, any federal land bank may establish branches within its district. Each district bank, before beginning business, must have a subscribed capital of not less than $750,000, divided into shares of $5 each. The capital may be increased from time to time. Stock subscriptions are required from national farm loan associations, and may be received from individuals, firms or corporations or from the government of any state, and, on organization of the bank, from the United States.

If within thirty days after the opening of books of subscription for the capital stock, any part of the $750,000 has not been subscribed, the Secretary of the Treasury shall subscribe the balance on behalf of the United States. While funds of the United States may thus be used in completing the minimum stock subscription on the organization of a federal land bank, provision is made for the retirement of the stock held by the United States when additional subscriptions are received from other sources. Government owned stock receives no dividends, but other stock shares in dividend distributions.

Each national farm loan association and the Government of the United States shall be entitled to one vote for each share of stock held by it in deciding all questions at meetings of stockholders, but no other shareholder is permitted to vote. Stock owned by the United States is to be voted by the Farm Loan Commissioner, as directed by the Federal Farm Loan Board.

Every federal land bank has power to borrow money, to give security therefor, and to pay interest thereon; to accept deposits of securities or current funds from national farm loan associations holding its stock, but to pay no interest thereon; to issue, subject to the approval of the Federal Farm Loan Board, and to sell farm loan bonds, which must be secured by first mortgages upon farm land deposited with the farm loan registrar for the district; to invest its funds in the purchase of first mortgages on farms in its district; to buy and sell United States bonds and to deposit its securities, and its current funds subject to check, with any member

Digitized by Google

bank of the Federal Reserve System, and to receive interest thereon. When necessary or convenient, a federal land bank may acquire and dispose of real or personal property, and may lease the same in part for revenue purposes, but, without the special approval of the Federal Farm Loan Board, may not retain for longer than five years possession of any real estate purchased or acquired to secure a debt.

No federal land bank has power to accept deposits of current funds payable upon demand except from its own stockholders, to make loans except on first mortgages through national farm loan associations, to issue bonds in excess of 20 times its capital or surplus, or to receive from any national farm loan association additional mortgages when the principal remaining unpaid upon mortgages already received from such association exceeds 20 times the amount of its capital stock owned by such association.

Bonds issued by federal land banks may be in denominations of $25, $50, $100, $500 and $1,000, in series of not less than $50,000. They shall have interest coupons attached, payable semi annually, and the rate of interest must not exceed 5 per cent. per annum.

All bonds are issued against collateral security of United States Government bonds, or first mortgages on farm lands, at least equal in amount to the bonds issued. The Federal Farm Loan Board may, at any time, call upon any land bank for additional security to protect the bonds it has issued. The security is made still stronger by a provision that when any farm loan bank shall default in paying the interest coupons on its bonds, the other such banks shall be liable to pay them in equitable proportion, and shall be liable also to pay proportionately the unpaid principal of farm loan bonds issued by any such bank which shall not be paid after the assets of the bank shall have been liquidated and distributed.

Not less than 25 per cent. of the capital of any land bank for which stock is outstanding in the name of national farm loan associations must be held in readily marketable securities or other liquid assets. Five per cent. of the capital must be invested in United States Government bonds.

Digitized by Google

Each federal land bank temporarily must have five directors and, when the subscriptions of national farm loan associations to its stock have reached $100,000, nine directors, of whom six, called local directors, are nominated and elected by the directors of the national farm loan associations in the district. These six local directors designate two of their number to serve one year, two to serve two years, and two to serve three years, their successors to be elected for terms of three years. The other three directors, called district directors, are appointed by the Federal Farm Loan Board and represent the public interest. In case of a tie the Farm Loan Commissioner determines the choice. The Federal Farm Loan Board designates one of the district directors to serve for three years and to act as chairman, one to serve for two years, and one for one year, their successors to be appointed for terms of three years. A director must have been a resident of the district for at least two years, and during his term of office may not engage in any other banking or mortgage loan business. At least one district director shall be experienced in farming and be actually engaged in farming in the district.

The board of directors, subject to the approval of the Federal Farm Loan Board, fix the salaries of the directors, officers, or employees, to be paid from the general funds of the federal land bank.

National Farm Loan Associations. In any farm loan district 10 or more natural persons may unite to form a corporation known as a national farm loan association, which shall be composed only of those who are owners or about to become owners of farm lands and of those who are borrowers or about to become borrowers on such land. An application to the federal land bank for a charter must be accompanied by a subscription to the capital stock of the bank equal to 5 per cent. of the aggregate sum desired on mortgage loans, which sum may not be less than $20,000.

Upon receipt of the stock subscriptions and of articles of association, the federal land bank sends an appraiser to investigate the solvency and character of the applicants for a charter, and then determines whether a charter for the

Digitized by Google

association should be granted. The recommendation of the land bank is forwarded to the Federal Farm Loan Board. If it is unfavorable, a charter is refused; if favorable, the board grants a charter, unless good cause is shown why this should not be done, and designates the territory in which the association may make loans. Upon receipt of its charter, such national farm association is empowered to receive from the federal land bank of the district sums to be loaned to its members on first mortgage farm land security.

A national farm loan association is organized on a coöperative basis. Its capital stock, consisting of shares of a par value of $5 each, may be increased from time to time for the purpose of securing additional loans for members. Each shareholder participates ratably in dividends earned and distributed, and is entitled to one vote on each share of stock held by him at all elections of directors and in deciding all questions at meetings of shareholders, but in no case to more than twenty votes. Shareholders are individually and ratably held responsible for all debts and obligations of the association to the extent of the par value of the stock owned by them, in addition to the amount paid in and represented by their shares.

Every national farm loan association has power to indorse, and thereby to become liable for the payment of first mortgages accepted from its shareholders by the federal land bank of its district to secure loans on farm lands, and to issue certificates of deposits of current funds, bearing interest for not longer than one year, at not to exceed 4 per cent. per annum, convertable into farm loan bonds when presented at the district land bank. Such deposits must forthwith be transmitted to the bank, and invested by it in first mortgages or in the purchase of bonds issued by a federal land bank.

Every national farm loan association must, out of its earnings, semi annually carry to reserve account a sum not less than 10 per cent. of the net earnings until there is a reserve of 20 per cent. of the outstanding stock issue, and thereafter 2 per cent. of the net earnings is to be added annually to this account. After the deduction of this reserve any association may declare a dividend to shareholders of the

Digitized by Google

whole or any part of the balance of its net earnings. The reserve is to be invested under rules prescribed by the Federal Farm Land Board.

Any national farm loan association may make application to the land bank of its district for loans not exceeding in the aggregate one fourth of its total stock holdings in the bank, and the bank may make the loan and charge interest at not exceeding 6 per cent. per annum.

Every national farm loan association shall elect a board of not less than five directors, who elect a president, a vice-president, a secretary-treasurer, and a loan committee of three members. All the directors and officers except the secretary-treasurer must be shareholders in the association and residents in its territory.

The directors and all officers of the association, except the secretary-treasurer, serve without compensation unless the payment of salaries to them shall be approved by the Federal Farm Land Board. The salary of the secretary treasurer, fixed by the board of directors, and the reasonable expenses of his office and of the loan committee and other officers are to be paid from the general funds of the association.

The secretary-treasurer must furnish a surety bond, and it is his duty to act as custodian of the funds of the association and to deposit the same in such bank as the board of directors may designate, to pay over to borrowers sums received for their accounts from the federal land bank, to collect, receipt for, and transmit to the federal land bank payments of interest, amortization installments, or principal arising from farm mortgage loans, and to be custodian of the stock certificates, securities, and records of the association. He must also make a report to the Federal Farm Loan Board, on specified forms, quarterly and when requested by the board. From time to time he is to assure himself that the loans made through the association are applied as set forth in the applications for loans by borrowers. He must report to the district land bank the failure of any borrower to comply with the terms of his mortgage or to pay taxes on the property mortgaged.

Method of Obtaining Loans. Loans made through national farm loan associations to its members by federal land banks,

Digitized by Google

may not be for less than $100 nor more than $10,000 to any one borrower, and not for less than 5 years nor more than 40 years. Loans in any bank district must be secured by first mortgages on farm land within the district, and must not exceed 50 per cent. of the value of the land mortgaged and 20 per cent. of the value of the permanent, insured improvements thereon, the value to be ascertained by appraisal, the basis of appraisal to be the earning power of the land for agricultural purposes. No loan may be made to any person who is not at the time, or shortly to become, engaged in the cultivation of the land mortgaged. The par value of shares of stock in national farm loan associations is $5 each, and every borrower must own one share for each $100 borrowed. In case of the sale of the mortgaged land, the federal land bank may permit the mortgage and stock interest of the vendor to be assumed by the purchaser. In case of the death of the mortgagor, his heirs or legal representatives have the option, within 60 days, of assuming the mortgage and stock interests of the deceased.

Loans may be made for the following purposes only:

To purchase land for agricultural uses.

To purchase the equipment, fertilizers, and live stock necessary for the proper and reasonable operation of the mortgaged farm, and to provide buildings for the improvement of farm lands, the terms "equipment" and "improvement" to be defined by the Federal Farm Loan Board.

To liquidate indebtedness of the owner of the land mortgaged incurred for the foregoing mentioned purposes.

Any person desiring to secure a loan must make application for membership in a national farm loan association, and state, on a prescribed form, the objects to which the proceeds of the loan are to be applied. He must also subscribe for one share of the association's capital stock for each $100, or major fractional part thereof, of the amount that he desires to borrow. The value of these shares, and the cost of appraisal, legal fees and recording charges are paid by the borrower, or, by arrangement with the federal land bank lending the money, may be included in the loan and deducted from the amount of the loan, provided that the whole amount does not exceed

Digitized by Google

$10,000. The applicant may become a member of the association by a two thirds vote of its directors. If the loan be granted he, on payment for the shares, becomes the owner of them. They are held by the association as collateral security for the payment of the loan, but the borrower receives the dividends on them while outstanding.

The application for a loan is referred to the loan committee of the association, composed of three members, who examine the land offered as security and make a detailed report. No loan may be approved by the association's directors unless the committee's report is favorable. This report is submitted to the federal land bank of the district, together with the application for the loan. The bank then refers the application and the loan committee's report to one or more of the land bank's appraisers, who investigate and make a written report upon the land offered as security, and no loan may be made unless this report is favorable.

Whenever any national farm association shall desire to secure for any member a loan on first mortgage from the federal land bank of its district, it must subscribe for capital stock of the bank to the amount of 5 per cent. of the desired loan, such subscription to be paid in cash upon the loan being granted by the bank. Such shares of stock are held by the bank as collateral security for the payment of the loan, but the association is paid any dividends accruing on such stock outstanding.

Every mortgage which a borrower makes to his association to secure his loan must be indorsed by the association to the bank that grants the loan, the association thereby making itself liable for the payment of the mortgage. Every such mortgage must be on an amortization plan, by means of a fixed number of annual or semi annual installments sufficient to cover, first, a charge on the loan, at a rate not exceeding the interest rate in the last series of farm loan bonds issued by the bank granting the loan; second, a charge for administration and profits at a rate not exceeding 1 per cent. per annum on the unpaid principal, said two rates combined constituting the interest rate on the loan; and, third, such amounts to be applied on the principal as will extinguish the debt within an

Digitized by Google

agreed period, not less than 5 nor more than 40 years. Before the first issue of bonds by any bank the rate may be determined by the bank, but at no time may the interest rate on loans exceed 6 per cent. per annum, exclusive of amortization payments. Interest on defaulted payments, however, may be at the rate of 8 per cent.

Any association is allowed to retain as a commission from each interest payment on any loan indorsed by it an amount to be determined by the Federal Farm Loan Board, not to exceed one eighth of one per cent. semi annually upon the unpaid principal of the loan, any amounts so retained to be deducted from dividends payable to the association by the federal land bank.

If the borrower expends the whole or any portion of a loan for purposes other than the specified objects for which farm loans are allowed, or if he defaults in any covenant of the mortgage, the whole of the loan may, at the option of the mortgagee, become due and payable forthwith. Taxes, liens, judgments, or assessments on the mortgaged land which are paid by the mortgagee become a part of the debt and bear simple interest at 8 per cent. per annum. Every borrower must undertake to keep insured all buildings the value of which was a factor in determining the amount of his loan. Insurance must be made payable to the mortgagee, as its interest may appear at the time of loss, and at the option of the mortgagor, subject to general regulations of the Federal Farm Loan Board, sums so received may be used for reconstruction of the buildings destroyed.

In the discretion of the federal land bank making a loan, a reappraisal of the mortgaged land may be permitted at any time, and such additional loan may be granted as such reappraisal will warrant.

After five years from the date of a loan, payments in addition to amortization payments may be made in sums of $25 or any multiple thereof for the reduction of the principal, or payment of the entire loan may be made, on any regular installment date.

The shares of stock in a federal land bank paid for by a national farm loan association when it secures a loan for

Digitized by Google

one of its members from the bank may, in the discretion of the bank's directors, and with the approval of the Federal Farm Loan Board, be paid off at par and retired, and must be paid off and retired upon full payment of the mortgage loan. In such cases the national farm loan association pays off and retires the corresponding shares of its stock which were issued to the mortgagor when the loan was secured from the bank.

Whenever it shall appear to the Federal Farm Loan Board that national farm loan associations have not been formed and are not likely to be formed in any locality, because of peculiar local conditions, the board may authorize federal land banks to make loans on farm lands through agents approved by the board. No agent may be employed, however, other than an incorporated bank, trust company, mortgage company, or savings institution chartered by the state in which it has its principal office.

Loans made by such agents are subject to the same conditions and restrictions as if they were made through national farm loan associations, and each borrower must become a shareholder in the federal land bank granting the loan to the extent of 5 per cent. of the amount he borrows.

Any agent negotiating such a loan must indorse the same and become liable for the payment thereof on default by the mortgagor. When required by the federal land bank, the agent collects and forwards to it without charge all interest and amortization payments on loans the agent indorses. The bank may allow the agent the actual expenses of appraising the land offered on security, of examining and certifying the title, and of making, executing and recording the mortgage papers, and in addition one half of 1 per cent. per annum upon the unpaid principal of the loan, such commission to be deducted from dividends payable to the borrower on his stock in the bank.

If there shall be a default under the terms of any indorsed mortgage held by a federal land bank, the national farm loan association or agent through which the mortgage was received by the bank is notified of the default, and the association or agent is required within 30 days to make good the

Digitized by Google

default, either by payment in cash of the amount unpaid, or by the substitution of an equal amount of the bank's farm loan bonds, with all unmatured coupons attached.

Joint Stock Land Banks. In some localities farmers may dislike the publicity and scrutiny which are necessary in obtaining loans through national farm loan associations from a federal farm loan bank. On account of objections to coöperative undertakings, such associations may be formed slowly in some states, with the result that farmers in a wide field could not secure credit accommodation at reasonable rates.

The act, therefore, provides for the organization of joint stock land banks, with private stock subscriptions, for the purpose of lending on farm mortgage security and issuing farm loan bonds. These banks can not engage in other business or receive deposits. The United States Government will not be a shareholder, as it may be in the organization of a federal land bank. Any number of natural persons not less than ten may organize a joint stock land bank, which must have not less than five directors. Share holders are individually and ratably held responsible for all the contracts and debts of such bank to the extent of the face value of the stock they own, in addition to the amount paid for their shares.

A joint stock land bank may be chartered by the Federal Farm Loan Board when it has capital stock to the amount of at least $250,000 subscribed, and it may begin business when one half of that amount is paid in and the balance is subject to call by the board of directors. It may not issue bonds, however, until the stock subscriptions are entirely paid, and it may not issue bonds in excess of 15 times the amount of its capital and surplus.

Joint stock land banks are restricted to a charge of 6 per cent. on farm loans, as in the case of federal farm banks, and are not allowed to charge a rate of interest exceeding by more than 1 per cent. the rate of interest on their last bond issue, and they may in no case demand or receive, under any form or pretence, any commission or charge not specifically authorized by the federal farm loan act. No joint stock bank may make loans not secured by first mortgages on farm

Digitized by Google

lands within the state in which is its principal office, or within some one state contiguous to that state.

Certain restrictions on the loans of federal farm loan banks do not apply to joint stock land banks, such as limiting the amount of loans and requiring that loans be made to farmers only and that they be used for only specific purposes. The loans of joint stock land banks must be payable in amortization installments, and in other respects so far as applicable such banks are subject to the restrictions and conditions imposed on federal farm land banks.

The Federal Farm Loan Board appoints land bank examiners, who are subject to the same requirements, responsibilities, and penalties as are applicable to national bank examiners under the Federal Reserve act. They examine federal land banks and joint stock land banks at least twice each year, and national farm loan associations whenever so directed by the board.

Every federal land bank, and every joint stock land bank must semi annually carry to reserve account 25 per cent. of its net earnings, until there is a reserve of 20 per cent. of the bank's outstanding stock issue, and thereafter 5 per cent. of the bank's earnings are to be added annually to this account. After the deduction of this reserve every federal land bank or joint stock land bank may declare dividends to its shareholders of the whole or any part of the balance of its net earnings. The reserves are to be invested under rules prescribed by the Federal Farm Loan Board.

The capital and reserve or surplus of every federal land bank and of every national farm association, and the income derived therefrom, are exempt from national, state, municipal, and local taxation. Mortgages held by federal land banks and joint stock land banks to secure loans on farm lands, and farm loan bonds issued by such banks and the income derived from these mortgages and bonds are also exempt. The shares of stock in federal land banks are exempt, but under the terms of the act the shares of joint stock land banks are not exempted, and real estate however acquired by either federal land banks or joint stock land banks is not exempted.

Digitized by Google

Farm loan bonds issued by federal land banks or joint stock land banks are a lawful investment for all fiduciary and trust funds, and may be accepted as security for all public deposits. Such bonds may be bought or sold by any member bank of the Federal Reserve System.

When so designated by the Secretary of the Treasury, federal land banks and joint stock land banks may be depositories for public money, except receipts from customs, but no Government deposits may be invested in mortgage loans or farm loan bonds.

Digitized by Google

WAGES IN AMERICAN CITIES.

By W. F. Ogburn and Esther Kelly, *Reed College, Portland, Oregon.*

Studies in the comparative statistics of cities are of importance because they place emphasis on extending the realm of knowledge about our cities and because they stimulate a healthy rivalry. Readers will appreciate the value of such studies by recalling comparative statistics with which they are familiar, such as death rates, figures of illiteracy, percentages of street area paved, and growth of population. The purpose of this paper is to direct attention to a possible ranking of cities in the average amount of wages paid. The importance of knowing wage conditions is of course very great and is emphasized particularly when we remember that we are living in an age when the fight against poverty is greater than ever before, when we are setting a minimum wage in one state after another and when many investigators are claiming that low wages are the one factor to which an increasingly large number of social ills can be traced.

In Table I is seen the ranking of 37 cities of the United States in the average union wage per hour in 10 trades. The data from which the averages and ranking were computed are for May, 1914, published in Bulletin No. 171 of the United States Bureau of Labor Statistics. It would be desirable to have a larger list of cities but the range of choice was unfortunately limited to the 41 cities from which the Bureau collected statistics. The list of 41 cities was reduced to 37 in number because data for the 10 trades were not given in 4 of the cities. These cities, however, are located in 29 different states and are the largest cities in their respective sections. They also represent typical geographical divisions and typical industrial areas, such as coal regions and textile districts.

The ranking of the cities is on the basis of average hourly union wages in the following 10 trades: bricklayers, carpenters, painters, plasterers, plumbers and gas fitters, day compositors in book and job printing, night compositors in book and job

Digitized by Google

printing, day compositors in newspaper printing, day and night linotype operators. It would have been desirable to include wages for unskilled work and salaries, but these were impossible to obtain. Indeed it was not possible to obtain data for all of the 37 cities in more than these 10 trades. For the different cities data were given for 20 to 80 trades but only 10 trades were found common to all of the 37 cities.

TABLE I.

RANKING OF 37 CITIES IN AVERAGE UNION WAGES PAID IN 10 TRADES. (a)

Rank.	Cities.	Cents per Hour.
1	Seattle	71
2	San Francisco	69
3	Portland, Ore.	67
4	Salt Lake	65
5	Denver	64
6	Chicago	64
7	St. Louis	64
8	New York	64
9	Los Angeles	62
10	Kansas City, Mo.	62
11	Pittsburgh	61
12	Boston	60
13	Newark	59
14	Washington	58
15	Cincinnati	58
16	St. Paul	57
17	Memphis	57
18	Cleveland	57
19	Omaha	56
20	Indianapolis	56
21	Detroit	56
22	Birmingham	55
23	Milwaukee	54
24	Little Rock	54
25	Louisville	54
26	New Orleans	53
27	Buffalo	53
28	Springfield, Ill.	53
29	Baltimore	53
30	Jacksonville	53
31	Providence	52
32	Minneapolis	52
33	Scranton	50
34	Philadelphia	50
35	New Haven	50
36	Atlanta	43
37	Charleston, S. C.	38

(a) The 10 trades are bricklayers, carpenters, painters, plasterers, plumbers, gas fitters, day compositors in book and job printing, night compositors in book and job printing, day compositors in newspaper printing, and day and night linotype operators.

The ranking of the cities in Table I may be considered as accurate for the 10 trades. But we are interested in knowing the ranking of these cities in all trades, as well as in a selected list of 10 trades. The question then is: How satisfactory an index is the ranking in wages in 10 trades for the

Digitized by Google

ranking in wages in all trades? How true is the sample of the whole? Since data for all trades or a very large number of trades are not given for every one of the 37 cities, how can we test the truth of the ranking in 10 trades? Although wages in a very large number are not given for every one of the 37 cities, how can we test the truth of the ranking in 10 trades? Although wages in a very large number of trades are not given for every one of the 37 cities, a very good approximation can be made by resorting to the device of breaking up the list of 37 cities into small groups. For the cities of each small group, there will be a greater number of common trades than there was for the total number of cities. A ranking may then be made for the cities in each small group on the basis of a large number of trades and comparisons made with the ranking of the same cities on the basis of the 10 trades. Upon comparison, if the relative position of any given city remains practically the same in both rankings, then the ranking on the basis of 10 trades may be accepted as a good sample, a satisfactory index of the whole. The geographical divisions were found most convenient for separating the cities into small groups and a larger number of common trades were actually found. For example in the column on the left below are listed the 6 Western cities and the average wage paid in the 10 trades; the figures in the right-hand column represent the average wages in the 35 trades that were found common to the cities of this small group.

Seattle	71	San Francisco	66
San Francisco	69	Seattle	64
Portland	67	Portland	61
Salt Lake	65	Salt Lake	59
Denver	64	Denver	57
Los Angeles	62	Los Angeles	57

The only change in the ranking is in the interchange of Seattle and San Francisco in the first place. For the Western cities, then, the ranking on the basis of 10 trades may be considered as a good index of a ranking on a large number of trades.

Digitized by Google

In Table II is found the comparison in the ranking of all of the cities by groups, in the manner presented above for the Western cities.

TABLE II.

COMPARISON OF RANKING OF CITIES IN WAGES PAID IN 10 TRADES WITH RANKING IN WAGES PAID IN A LARGER NUMBER OF TRADES.

Groups of Cities and Number of Trades in Each.		Average Hourly Union Wage Paid in 10 Trades.	Average Hourly Union Wage Paid in Large Number of Trades.	Rank on Basis of 10 Trades.	Rank on Basis of Larger Number of Trades.
North Atlantic, 20	New York	64	63	1	2
	Pittsburgh	61	55	2	4
	Boston	60	59	3	3
	Newark	59	63	4	1
	Buffalo	53	51	5	6
	Providence	52	53	6	5
	Scranton	50	49	7	9
	Philadelphia	50	50	8	7
	New Haven	50	50	9	8
South Atlantic, 13	Washington	58	57	1	1
	Baltimore	53	51	2	2
	Jacksonville	53	52	3	3
	Atlanta	43	43	4	4
	Charleston	38	34	5	5
North Central, 20	Chicago	64	62	1	1
	St. Louis	64	61	2	2
	Kansas City	62	60	3	3
	Cincinnati	58	54	4	8
	St. Paul	57	53	5	10
	Cleveland	57	55	6	4
	Omaha	56	54	7	7
	Indianapolis	56	53	8	9
	Detroit	56	54	9	6
	Milwaukee	54	50	10	12
	Springfield	53	51	11	11
	Minneapolis	52	54	12	5
South Central, 16	Memphis	57	54	1	3
	Birmingham	55	55	2	2
	Little Rock	54	56	3	1
	Louisville	54	51	4	5
	New Orleans	53	54	5	4
Western, 35	Seattle	71	64	1	2
	San Francisco	69	66	2	1
	Portland	67	61	3	3
	Salt Lake	65	59	4	4
	Denver	64	57	5	5
	Los Angeles	62	57	6	6

Columns five and six show the ranking of the cities in wages on the basis of 10 trades and on the basis of a larger number of trades. Comparing the figures in these two columns it is seen that the rank of one half of the cities is the same on the basis of 10 trades as when the wages for a larger number are used and that, although one city is shifted seven places in the ranking, the average shift is one place. It is true that the

Digitized by Google

number of trades is small in the second ranking and even greater changes would be found if the whole list of cities served rather than the small groups as the basis for the ranking. The conclusion of this somewhat detailed analysis is that on the average a city in Table I is not removed by more than two or three positions from the rank it would have received in a table constructed on the basis of a larger number of trades.

One of the most striking points to be noted in Table I is the relatively low ranking of Philadelphia as compared with New York. This difference in wages paid in two large port cities in the same economic section of the country and separated by such a short distance was so unexpected that it seemed desirable to test the results further. Accordingly the two cities were compared on the basis of hourly union wages paid in 69 common trades, with a result which showed that in New York an average wage of 54 cents an hour was paid, while in Philadelphia for the same 69 trades, an average hourly wage of 43 cents was paid. Thus, using the wage for 69 common trades, the average hourly wage in New York was found to be 11 cents higher than in Philadelphia, whereas on the basis of 10 trades the difference was found to be 14 cents. The difference in the average wages paid in the two cities is slightly lowered when a larger number of trades is considered, a result which would doubtless be found to hold with all of the cities if they were similarly tested. Therefore, judging from these added data, there is no reason to think that the relative ranking of New York and Philadelphia in Table I is not approximately correct.

It should be remembered that the data in Table I show not the actual wages paid, but rather only the ranking of the cities. If there were a very large number of union trades common to all of the cities, the actual wages paid would probably be some 10 cents an hour less than the figures given in Table I. For we found that in New York the average wage of 69 trades is 10 cents less than the average wage of 10 trades and in Philadelphia the average wage of 69 trades is 7 cents less than the average wage of 10 trades.

Digitized by Google

HIGHWAY ACCIDENTS IN NEW YORK CITY DURING 1915.

By Mills E. Case, *Secretary to the Police Department, City of New York.*

The subject of highway accidents must receive a great deal of attention in the future. At the present time, the number of persons killed or injured annually in the streets of our cities is very large, but the general public is not aware of the fact. In New York City, during 1915, the fatalities from street accidents were 2½ times the number of persons murdered in the entire city, yet it is probable that the press of the city gave 10 times as much space to murders as to accidents. Unless radical measures are taken the number of accidents will increase rapidly in all large American cities, especially in the older cities which have narrow streets. The steel skeleton type of construction has made possible skyscraper office buildings and apartment houses. The old streets are inadequate for the population of the new buildings. Some relief from street congestion is obtained through rapid transit lines, especially subways. Even the dangerous motor vehicle has done something to reduce street congestion. It takes less space on the street than the vehicles and horses which it supplants; it goes faster and so occupies the space a shorter time; it carries a larger load and so makes fewer units necessary. But each successive year a larger fleet of vehicles and a larger crowd of pedestrians must use the streets; each year the resident population increases; each year the number of visitors increases; each year more women "go to business."

In the congested residential sections of our large cities, the children must play in the streets. The old fashioned house had a yard; the tenements and apartment houses of today cover practically the whole of the lot and frequently contain children enough to fill a play-ground larger than the bare lot would be. This is a condition which is getting worse every year.

Street accidents in New York City have been recorded for several years in the precinct "Arrest and Aided Record,"

Digitized by Google

but I cannot find that any compilation of accident statistics was made prior to 1914. Police Commissioner Arthur Woods saw the importance of the matter and ordered a compilation of the statistics from the precinct records for the calendar year. He also authorized a system of special reports on highway accidents, which went into effect January 1, 1915. These reports are forwarded to the Bureau of Statistics at Headquarters, where they are compiled monthly in very elaborate tables.

Prior to August, 1915, the statistics were compiled by hand; since then, they have been compiled by means of perforated cards, with the necessary machinery for handling them. During the earlier months, it was not practicable to make the detailed studies which the mechanical method rendered possible in the later months. The tables for the year, presented herewith, are necessarily limited to the extent of the early tables. A word of caution is needed concerning the comparability of these tables with the 1914 tables (published in the Annual Report of the Police Department, 1914). Because of the supervision of the Bureau of Statistics, the 1915 statistics are more nearly complete than those for 1914. An effort was made during 1915 to obtain and include the after-mortality, of which there is no record in earlier years.

The tables must be read in the light of the following limitations upon the terms used:

Accident—In these tables no accident is included unless a person is killed or injured. The records and the reports to the Bureau of Statistics include all highway accidents known to the police. The elimination of the accidents in which no person is injured leaves a fair "sample" for administrative purposes. About 20 per cent. of the reports received are eliminated from the tabulation, because no one was injured.

Highway—Includes bridges, paths in parks; excludes elevated or subway trains, stations and stairs, ferries and other boats and piers or docks which are not a part of the highway. Some accidents which occur in the highway are excluded because independent of the place of occurrence, *e. g.*, falls on sidewalk due to intoxication, illness, or play. About 5 per cent. of all reports are excluded on this ground. The eliminating

Digitized by Google

is done at the Bureau of Statistics and the precinct officers are encouraged to forward all doubtful cases.

Fatal—The 1915 tables are corrected for after-mortality reported up to January 31, 1916. The correction began to be applied in February, 1915, hence, some January fatalities are not included, because unknown. No deaths from 1914 accidents are included in these tables.

Collision—During the first six months of 1915, this term includes cases of bicycles struck by a heavier vehicle. Since then, such cases are included with "Persons Struck By—."

Cause of Accident—There is an annoying overlapping of categories here due to the fact that the original reports indicate the causes in terms of varying generality. It is not practicable to return the reports for amendment in this respect. Rather than lose the specific information by throwing it into the general classes, the Bureau has preferred to use the awkward grouping which fits the material.

Digitized by Google

HIGHWAY ACCIDENTS IN WHICH PERSONS WERE INJURED OR KILLED IN THE STREETS, SIDEWALKS, PUBLIC HIGHWAYS, OR ON THE ROADS AND WALKS IN PUBLIC PARKS, NEW YORK CITY, 1915.

Nature of Accidents.	Accidents.			Persons Killed.					Persons Injured.				
	Total.	Fatal.	Non-Fatal.	Total.	Male.	Female.	Under 6 Years.	6 Years to 16 Years.	Total.	Male.	Female.	Under 6 Years.	6 Years to 16 Years.
Persons struck by—													
Railway trains	23	13	10	13	11	2	1	2	11	8	3	1	2
Street cars	1,771	77	1,694	77	58	19	8	10	1,730	1,227	503	277	368
Motor vehicles—													
Passenger	4,865	177	4,688	178	127	51	27	72	4,830	3,733	1,097	537	1,742
Truck or delivery	1,212	105	1,107	105	82	23	15	43	1,136	885	251	144	398
Motorcycles	377	...	377	...	...	...	...	...	414	270	144	63	114
Horsedrawn vehicles—													
Passenger	249	3	246	3	3	...	1	...	252	167	85	77	92
Truck or wagon	2,441	100	2,341	100	75	25	42	37	2,379	1,772	607	629	766
Saddle horse	16	...	16	...	...	...	...	...	16	13	3	4	3
Bicycles	352	3	349	3	2	1	...	1	357	226	131	71	158
Other vehicles	72	1	71	1	1	...	...	...	71	51	20	17	26
Falls and miscellaneous injuries received while riding on, boarding or alighting from vehicles	3,026	47	2,979	47	45	2	...	1	3,026	2,230	796	33	149
Falls, etc., while stealing rides	313	9	304	9	9	...	1	7	305	296	9	46	212
Collisions	2,428	41	2,387	42	38	4	2	5	2,982	2,489	493	46	202
Other accidents to vehicles	573	19	554	21	17	4	...	...	779	636	143	10	44
Runaways	421	13	408	13	9	4	5	1	483	431	52	23	30
Non-vehicular accidents—													
Falls	3,005	24	2,981	30	20	10	...	2	3,093	1,811	1,282	190	411
Hit by falling objects	584	1	583	1	...	1	...	1	624	436	188	48	108
Bitten by dogs	120	...	120	...	...	...	...	...	125	101	24	26	64
Miscellaneous accidents	692	16	676	16	13	3	4	2	708	584	124	59	196
Totals	22,540	649	21,891	659	510	149	106	184	23,321	17,366	5,955	2,301	5,085

Digitized by Google

ACCIDENTS BY MONTHS.

Nature of Accidents.	Total.	January.	February.	March.	April.	May.	June.	July.	August.	September.	October.	November.	December.
Persons struck by—													
Railway trains	23	1	1	1	...	2	2	2	3	3	1	3	4
Street cars	1,771	116	152	184	174	140	145	151	157	149	151	125	127
Motor vehicles—													
Passenger	4,865	272	211	289	396	432	424	501	438	582	610	431	277
Truck or delivery	1,212	100	69	99	86	92	88	137	119	104	131	91	96
Motorcycle	377	7	15	23	38	46	48	47	47	40	35	22	9
Horse-drawn—													
Passenger	249	11	11	24	27	28	11	31	28	30	26	14	8
Truck or wagon	2,441	123	136	204	247	225	225	261	242	230	242	177	129
Saddle horse	16	2	1	1	2	2	...	3	...	4	1	...	...
Bicycles	352	6	13	18	39	50	47	59	45	40	23	12	...
Other vehicles	72	...	...	6	2	6	2	22	5	11	6	7	5
Falls and miscellaneous injuries received while riding on, boarding, or alighting from vehicles	3,026	192	155	202	230	300	326	363	339	323	225	184	187
Falls, etc., while stealing rides	313	9	8	22	22	27	40	34	42	42	41	18	8
Collisions	2,428	137	161	176	189	241	285	229	201	197	209	211	192
Other accidents to vehicles	573	23	26	41	39	51	49	53	53	58	69	61	50
Runaways	421	25	22	40	40	38	46	50	33	25	34	34	34
Non-vehicular accidents—													
Falls	3,005	387	263	300	297	201	209	207	204	187	204	195	351
Hit by falling objects	584	41	33	41	64	67	73	58	61	46	31	36	33
Bitten by dogs	120	6	4	6	11	15	22	16	6	11	10	8	5
Miscellaneous accidents	692	122	78	48	51	78	76	67	48	44	35	19	26
Total	22,540	1,580	1,359	1,725	1,956	2,041	2,118	2,291	2,071	2,126	2,064	1,648	1,541

Digitized by Google

VEHICULAR ACCIDENTS.

Cause.	Number of Accidents.	Cause.	Number of Accidents.
Fault or incapacity of injured person—		Leaving horse unattended	14
Stealing ride	417	Intoxication	38
Improperly riding on vehicle	208	Illness	2
Improperly boarding or alighting	1,290	Carelessness	105
Falling under or against vehicle	147	Other faults	10
Running into or in front of vehicle	955	**Defects of vehicles—**	
Walking, standing or sitting in street or on curb	220	Steering gear	76
Crossing street not at crossing	3,289	Brakes	43
Skating or playing ball in street	172	Wheels	47
Playing in street	1,099	Axle or fifth wheel	50
Coasting or riding toy vehicle in street	32	Harness, pole, etc.	61
Improperly riding bicycle	52	Unmanageable vehicle	79
Intoxication	330	Vehicle starting or stopping suddenly	75
Illness	22	Other defects	43
Crippled condition	29	**Skidding of vehicle—**	
Blindness	3	Slippery street:	
Deafness	26	With tire chains	100
Working in street	150	Without tire chains	227
Carelessness	203	Other causes:	
Other faults	22	With tire chains	10
Fault of driver—		Without tire chains	80
Speeding	67	**Other causes—**	
Recklessness	85	Slippery streets	26
On wrong side of street	175	Excavation in street	2
Turning corner improperly	54	Obstruction in street	7
Coasting	1	Object in street	38
Violating 8-foot law	16	Street lights out	5
Improperly towing vehicle	8	Street poorly lighted	3
Failing to signal	14	Other street condition	41
Disregarding signal	5	Fog	3
Lights out	23	Rain	5
No flag or light on projection	1	Snow	2
Vehicle improperly loaded	12	High wind	8
On sidewalk	20	Traffic conditions	75
Driving into area closed to traffic	2	Miscellaneous causes	186
Obstructing traffic	1	**No cause given**	7,486
Backing	36		
Leaving auto unattended	11	**Total**	18,139

Digitized by Google

REGULARITY OF EMPLOYMENT—NEED OF STANDARDIZATION OF METHODS FOR MEASURING REGULARITY OF EMPLOYMENT.*

By N. I. Stone.

The ascertainment of annual earnings of industrial workers has been a fascinating will o' the wisp to the government statistician. The lack of information necessary for this purpose, for which the average factory pay roll is notorious, is sufficient to drive the statistician to despair. Employees are designated by arbitrary numbers which are given without distinction as to sex, occupation, method of compensation, etc. Number 25 in a garment factory may stand one week for a cutter, a highly skilled male worker, receiving $25 per week, and the following week, upon the discharge of the cutter, the same number may be given to a "finisher," a girl worker who has earned $7.88 by steady plugging at prevailing piece rates. There is nothing to indicate that number 25 of the former week and number 25 of the latter are two distinct individuals. Add to this the shifting of employees from shop to shop and it will be clear why it is practically impossible to trace the annual earnings of individual workers in most industries.

In 1913 the writer was engaged in an investigation of wages in the Dress and Waist Industry of New York City under the joint auspices of the Manufacturers' Association and the Union, with a view to standardization of piece rates paid in that industry. The pay rolls proved no exception to the general rule. The extreme fluctuations of employment, however, due to seasonal influences in that industry made some kind of an estimate of the annual earnings of the workers imperative. For this purpose the index number method was resorted to for the first time, as far as the writer is aware, in the measurement of wage fluctuations. The succeeding steps used in the process were as follows:

* Presented at the dinner of the American Statistical Association, New York City, May 5, 1916, in connection with the discussion of the subject of Standardization of Statistical Methods.

Digitized by Google

TABLE I.

FLUCTUATIONS OF EMPLOYMENT AND WAGES IN THE DRESS AND WAIST INDUSTRY FOR 1912.

Week.	Number of Employees.	Amount of Wages Paid Out.	Per Cent. (Busiest Week=100). (a)	
			Employees.	Wages.
1	14,550	$130,484	71	53
2	15,443	153,598	75	63
3	16,190	165,549	79	67
4	16,960	180,673	83	74
5	17,849	192,382	87	78
6	18,329	202,506	89	82
7	19,146	212,972	93	87
8	19,633	221,929	96	90
9	20,183	210,614	98	98
10	20,409	244,981	99	100
11	20,521	245,494	100	100
12	20,524	245,177	100	100
13	20,374	236,951	99	97
14	19,313	194,583	94	79
15	19,298	201,271	94	82
16	19,145	208,053	93	85
17	18,974	203,595	92	83
18	18,478	192,435	90	78
19	18,078	185,635	88	76
20	17,501	178,117	85	73
21	16,761	166,905	82	68
22	16,509	156,863	80	64
23	16,350	167,384	80	68
24	16,514	169,487	80	69
25	15,961	159,534	78	65
26	14,514	141,406	71	58
27	12,292	105,559	60	43
28	10,816	95,279	53	39
29	10,304	94,149	50	38
30	11,147	105,290	54	43
31	12,429	115,379	61	47
32	13,945	138,336	68	56
33	15,677	159,322	76	65
34	16,888	177,802	82	72
35	17,686	195,967	86	80
36	18,363	191,971	89	78
37	18,107	170,519	88	69
38	18,660	197,465	91	80
39	19,030	215,996	93	88
40	19,199	215,151	94	88
41	19,388	220,809	94	90
42	19,686	216,626	96	88
43	18,695	202,148	91	82
44	17,791	188,115	87	77
45	16,769	165,736	82	68
46	16,428	167,753	80	68
47	16,378	162,302	80	66
48	15,867	149,128	77	61
49	16,384	162,455	80	66
50	16,604	169,000	81	69
51	16,879	170,462	82	69
52	16,290	150,827	79	61
Average	17,100	178,887	83.3	73

(a) In the column for employees the busiest week means the week in which the maximum number were employed; in the column for wages it means the week in which the maximum amount was paid.

Digitized by Google

First, the earnings of individual employees during the busiest week of the year were ascertained for each factory under investigation. Next, the total number of workers employed and the total amount of wages paid out each week of the year by each of those factories were ascertained. The data were arranged in tabular form, each factory forming a vertical column of 52 lines, each line representing the total wages paid by that factory during that week. The wages paid each week by the 260 shops for which these data were obtained, were then added horizontally across the table, the total weekly wages for the industry as a whole thus forming an additional vertical column.

The next step was to express the weekly wage totals for the industry in terms of index numbers. The week showing the highest wage total (the eleventh week, see Table I) was regarded as the busiest week for the industry as a whole and the wage total for that week expressed as 100. The wages for each of the remaining 51 weeks were expressed as percentages of the wages of the busiest week, the lowest week (the twenty ninth) showing only 38 per cent. and the average for the year being 73 per cent. of the wages in the busiest week of the year.

The report of the investigation was published as Bulletin 146 of the United States Bureau of Labor Statistics, and Table I is reproduced from page 160 of that Bulletin.

The annual average index number of wages, equal to 73, was used as a basis in arriving at an estimate of the average annual earnings of workers in different occupations. Thus, if the earnings of a woman operator during the busiest week were equal to, say, $10.00, her average weekly earnings during the year would be 73 per cent. of that, or $7.30, and her annual earnings 52 times $7.30, or $376.60.

In a supplementary investigation last winter, Miss Juliet Stuart Poyntz had occasion to use the raw material of the investigation covered by Bulletin 146. In order to ascertain the extent of variation of employment in different types of shops, she applied to separate shops the method used by the writer for the industry as a whole. The percentages for the individual shops calculated under her direction are shown in the following table:

Digitized by Google

TABLE II.

FLUCTUATIONS OF EMPLOYMENT IN VARIOUS SHOPS AND GROUPS OF SHOPS IN THE DRESS AND WAIST INDUSTRY.

Per Cent. of Employment.	High Grade Shops.		Low Grade Shops.		Whole Industry.	
	Number of Shops.	Per Cent. of all High Grade Shops.	Number of Shops.	Per Cent. of all Low Grade Shops.	Number of Shops.	Per Cent. of all Shops.
Less than 20	..	..	..	..	..	..
20–24	..	..	..	..	..	..
25–29	1	1.0	2	1.2	3	1.0
30–34	2	2.1	4	2.4	6	2.3
35–39	4	4.2	1	0.6	5	2.1
40–44	5	5.2	19	11.2	24	9.0
45–49	8	8.3	27	16.5	35	13.5
50–54	12	12.4	37	21.9	49	18.4
55–59	20	20.3	18	10.8	38	14.2
60–64	21	21.7	25	14.8	46	17.3
65–69	13	13.4	17	10.0	30	11.3
70–74	7	7.2	14	8.3	21	7.9
75–79	4	4.2	4	2.3	8	3.0
Total	97	100.0	168	100.0	265	100.0

Average (simple)	55.6
Average (weighted)	58.2

As will be seen from the preceding table, out of 266 shops there were only three shops with an average of 73 per cent., the figure for the industry as a whole obtained by the writer, and only four shops had higher averages, ranging from 75 to 78 per cent. The percentages for the remaining shops ranged from 27 to 72. Averaging the percentages of the separate shops and groups of shops for the industry as a whole, Miss Poyntz obtained the figure of 55.6 per cent. as a simple average and 58.2 per cent. as a weighted average.

The discrepancy between the results obtained by the two methods was so striking as to suggest at first that an error had crept into the calculations of one or the other investigation. An analysis of the tabular material, however, disclosed the interesting fact that the figures of 73 per cent. and 58.2 per cent. respectively for the industry as a whole obtained by the two methods were each mathematically correct, the discrepancy being due to the difference in method of application of the index number principle.

Digitized by Google

An examination of the diagram showing fluctuations of employment in the industry (p. 169 of Bulletin 146) will disclose the fact that the fluctuations, while roughly synchronous in different groups of shops, are not entirely coincident; the same is true as to individual shops, that is to say, while all the shops are more busy in the spring than they are in the summer, the highest peak in the employment curve may be reached during the tenth week in one shop, the ninth week in another, the eighth in a third, the eleventh in a fourth, and so on. Again, the extent of fluctuation differs for various groups of shops, and still more so for individual shops. The combining of the wages of all the shops into one total for each week had the effect of smoothing out the sharp peaks in the curves for the individual shops, the coinciding of the peak of one shop and the depression of another having a mutually neutralizing effect. In other words, the combination of what might have been 260 separate curves into one tended to flatten out the curve representing the industry as a whole. The flatter the curve, the less the fluctuations, and the nearer, of course, it approaches a straight line representing 100 per cent.

To put it in less technical terms: If the industry were so organized that workers unemployed in their own shops could be transferred to other, more busy shops, so that the entire industry formed in this respect an integral unit, the method of combining the wages of all the shops into one total for the industry as a whole would have been entirely proper; but, as a matter of fact, there is no such mobility of labor in the industry during the season. At the end of the season, a considerable number of workers do leave their places in search of better positions in other shops; but, while the season lasts, very few workers leave their shops. This means that the regularity of employment for the bulk of the workers in the industry is measured not by the quantity of work available in the industry as a whole, but by that to be had in their respective shops; and it, therefore, seems that the summary of conditions in the industry as a whole can be correctly formed by arranging the annual percentages of the separate shops, as was done by Miss Poyntz and not by the method used by the writer.

Digitized by Google

This runs counter to accepted statistical standards according to which the averaging of percentages is wrong and only absolute numbers should be dealt with.

The fact that the method adopted by the writer has since been followed in investigations of other industries conducted by the Federal Bureau of Labor Statistics as shown in its reports in Bulletins 183 and 190 lends added interest to the question of what is the correct method of measuring regularity of employment.

It is hoped that this account may lead to further discussion of the subject and contribute to the standardization of methods in the measurement of one of the most important of economic phenomena,—the fluctuation of employment and its effect upon annual earnings, in regard to which there is an extreme paucity of accurate data.

Digitized by Google

REVIEWS AND NOTES.

The Mortality from Cancer throughout the World. Frederick L. Hoffman. Newark, N. J.: The Prudential Press. Pp. 826+xv.

The statistical method of eliminating error by the force of massing large volumes of data is excellently exemplified by Frederick L. Hoffman's monumental monograph on cancer; for he includes in his survey a review of statistical returns for some 26 per cent. of the total population of the world.

Vital statistics from South American cities, from African provinces, and from Balkan states at first, of course, inspire a certain amount of suspicion. Yet the paradox enunciated by F. A. Woods,—to the effect that the more faulty the original source of data the more reliable are any conclusions to which they lead in spite of such faults,—is essentially a sound one. I assume this to mean that if the probable error or the element of random deviation is large, the causes which nevertheless produce a significant correlation must be very important indeed. This of course assumes no *constant* error, and the criticism of cancer statistics is based on the suspicion that such a constant error does exist in the shape of a tendency to better diagnosis. Many statisticians, following King and Newsholme and the British Imperial Cancer Research workers, have on this ground doubted the alleged increase in this disease, but the evidence brought forward by Mr. Hoffman certainly goes far to substantiate his conviction that the increase is a real one. The combined cancer rate for a group of European and Australasian countries increased from 44.8 per 100,000 in 1881 to 90.4 in 1911. Neither the changes in age distribution nor the improvement in medical diagnosis which have taken place can reasonably be held to account for such an enormous difference. The reported death rate is low among primitive peoples and where statistics are available its recent incidence varies from a standardized death rate of between 50 and 60 for Japan and Italy to rates between 100 and 110 for Bavaria and Switzerland. In countries of the latter type its increase has now been checked. In other words the recorded cancer rate tends to approach 100 or thereabouts. All this would in general fit in fairly well with the explanation of better diagnosis, but the detailed differences as Hoffman shows point to something more.

Particularly valuable is his analysis of cancer in different countries by the part of the body affected. Thus cancer of the stomach, liver, and oesophagus (strictly inaccessible cancers) are higher in Uruguay and Japan than in England and Wales or the United States, and more than twice as high in Switzerland, Holland, Norway, and Bavaria as in the United States, while cancer of the skin (easy of diagnosis) is much higher in the United States and England and Wales than in any of the other countries mentioned above, which facts can not be explained as due to medical diagnosis. So cancer of the female generative organs (difficult of diagnosis) is nearly as

Digitized by Google

common in Japan as in England, while cancer of the breast (easy of diagnosis) is a little more than one tenth as frequent. Both English and German statistics are cited to show that in the increase of recorded cancer during the past quarter century cancer of the accessible organs has played a more important part than that of the inaccessible organs. Altogether the impression made upon the reviewer by the mass of Mr. Hoffman's data is that the geographical and historical differences in recorded cancer death rates do for the most part represent real and significant biological facts.

The limits of a review forbid even a glance at Mr. Hoffman's discussion of the somewhat puzzling occupational statistics of cancer, or at his widely sweeping review of the geographical incidence of cancer throughout the civilized world. He finds evidence for a general increase in cancer as one proceeds northward and southward from the equator. He finds no evidence for inheritance of cancer or for any parasitical or communicable origin. The main purpose of the book is, however, to call attention to the fact "that the menace of cancer throughout the civilized world is much more serious than has generally been assumed to be the case" and to lay a foundation for future statistical study of the fundamental problems involved.

The last 500 pages of the work which constitute the appendices will prove of the greatest value to future investigators. They include the various approved classifications of tumors, the forms used for cancer records and questionnaires by leading English and American authorities, copies of popular circulars on the control of cancer, and a mass of statistical material copied from original official reports. Occupational cancer statistics are given for England and Wales, Hungary, and the United States (Prudential Insurance Co.). There are 85 pages of insurance statistics (English and American) largely by age and by the affected organ, 163 pages of cancer statistics for American states and cities (largely by organs), and 193 pages of foreign statistics (mostly crude cancer death rates by years). Finally there is a bibliography which occupies 18 closely printed pages.

Altogether this is perhaps the most impressive monograph which the science of vital statistics has yet produced in the United States. We owe much to Mr. Hoffman for his industry and enthusiasm in its preparation and to The Prudential Company for the broad vision and the public spirit which made possible the production and the publication of so notable a volume.

C.-E. A. Winslow.

Yale Medical School.

Digitized by Google

ANTHROPOMETRIC STATISTICS OF CHILDREN—AGES SIX TO FORTY-EIGHT MONTHS.

About five years ago, the *Woman's Home Companion* inaugurated a novel plan to direct public attention to the fact that our children are less perfect, physically, than is generally assumed to be the case. In its inception this movement in favor of greater emphasis on *healthy* children and less emphasis on *good-looking* children was in the form of Better Babies Contests, which were held in various localities throughout the United States, usually under the auspices of some local organization, but with the coöperation of the Better Babies Bureau of the *Woman's Home Companion.*

In these Better Babies Contests the children were examined under three groups of questions: First, with reference to mental development; second, with reference to physical defects or impairments; and, third, with reference to physical development.

The different items in these main groups were given ratings. The total or maximum rating under the mental and developmental test was 200; under physical examination for defects or impairments, the total maximum rating was 700; and under physical measurements the total maximum rating was 100—making a grand total of 1000 points.*

Early in 1914 plans were made for the holding of a Better Babies Contest in Newark, N. J., in March of that year. This contest was under the joint auspices of *Contemporary,* a prominent local organization of women, and the State Pediatric Society. The instructions, score cards, literature, medals and certificates were in this case, as in the many previous contests, provided by the *Woman's Home Companion.*

Up to this time, the table of physical measurement standards with which the children at ages under five years were compared was based upon too few numbers to warrant its use for anything more than approximate averages. This first table, also, was based upon average measurements of both sexes combined, and the difficulties of the scorers were further increased by reason of the fact that the average measurements were published only for three-month intervals for ages under one year—6 months, 9 months, and 12 months—and for four-month intervals at ages one year and upward—16 months, 20 months, 24 months, 28 months, etc.

As a result of the first Better Babies Contest held in Newark, N. J., I proposed that the table of physical standards be thoroughly revised, that it be based upon larger numbers, that males and females be considered separately, and that the average measurements be published for single months of age. These improvements were recommended as theoretically desirable even before this contest was held, and the statistics of the 542 children, ages 6 to 36 months, entered in that contest, proved that only with the aid of such a table could the physical proportions of any given child be rated fairly.

* In a later, revised score card the ratings are grouped as follows: Test I. Mental and Developmental, maximum total points 200; Test II. Measurements, maximum total points, 100; Test III. Physical Examination—maximum total, 500 points; Test IV. Oral and Dental Examination, 115 points; Test V. Eye, Ear, Nose and Throat, 85 points.

Digitized by Google

The Better Babies Bureau, of the *Woman's Home Companion*, invited me to compile such a table, and this was done in 1914, from measurements furnished by that Bureau and derived from Better Babies Contests which had been held in 23 different States during the years 1913 and 1914, including altogether 3,448 sets of measurements, of which 1,843 were of males and 1,605, of females. This first table was limited to ages 6 to 42 months, both inclusive, and as the statistics were compiled by single months the numbers for any given age were comparatively small. In spite of this fact, however, the averages gave exceptionally smooth results.

During the year 1914 a new development of the central purpose of Better Babies Contests was fostered by the Council on Public Health and Instruction, of the American Medical Association, but under the title of "Baby Health Conferences." At these conferences score cards similar to those used at the Better Babies Contests are filled out, including physical measurements. The physical measurements at both contests and conferences are made in a uniform manner and in accordance with instructions which go with the score cards. During the year 1915 these score cards of the American Medical Association were forwarded to me and the data were compiled and added to those already available through the courtesy of the *Woman's Home Companion*. In the early part of 1916 I found that the aggregate number of sets of measurements available exceeded 10,000, and this was thought to be a sufficiently large number to give satisfactory results for the purpose of compiling a revised standard anthropometric table for children of ages 6 to 48 months, both inclusive. This table, a copy of which is presented herewith, is to be published by the Committee on Public Health Education among Women, of the American Medical Association, and used in connection with the Better Babies Conferences held under the auspices of this Committee throughout the United States. This table, and the preliminary one of 1914, are thought to be the first for this country constructed in just this way—that is, from healthy children, all measured according to uniform rules and, in the main, by physicians, nurses, or other specially instructed persons. These tables differ somewhat from those referred to by pediatrists, such as Dr. Emmet L. Holt in his book entitled "Diseases of Infancy and Childhood." Dr. Holt's statistics for the first year of life are based upon observations made upon 100 healthy nursing infants at the Nursery and Child's Hospital, New York City. These children—50 males and 50 females—were weighed daily during the period of observation, and the purpose was to ascertain the normal rate of growth, which was found to be steady and fairly consistent from the third day. There were, of course, individual variations, but the average measurements —weight, height, etc.—increased gradually and consistently from the third day to the end of the first year of life.

Dr. Holt states that, "comparatively few observations have been published upon the weight during the second to the fifth years of life." He had, however, about 372 personal observations, from which he concludes that the gain in weight is about 6 lbs. during the second year, 4½ lbs. during the third year, and 4 lbs. during the fifth year. His observations were also

Digitized by Google

to the effect that throughout this period the gain in weight of girls was at about the same ratio as that of boys, but that the girls remained on the average nearly one pound lighter.

In the accompanying table it is interesting to note that during the second year of life the average increase in weight for males was 5¼ lbs., and for females, 5⅛ lbs. The increase in height during the second year of life was, for males, 4⅜ inches, and for females, 4½ inches. The accompanying table also shows that the average weight of boys at the age of six months is 1¼ lbs. more than the average weight of girls at the same age. The average height of boys at the age of six months is 26½ inches, as compared with an average height for girls at the same age of 25⅞ inches. In other words, boys at the age of six months, according to this table, are ⅝ of an inch taller on an aver-

AVERAGE PHYSICAL MEASUREMENTS. (a)
MALES.

Age, Months.	No.	Weight, Pounds.	Height.	Circumference of Head.	Chest.	Abdomen.	Diameter of Chest. Lateral.	Ant.-Post.	Length of Arm.	Leg.
6	259	18	26½	17⅜	17⅜	17¼	5⅜	4⅜	10¼	11⅜
7	256	19¼	27¼	17⅜	17⅝	17⅜	5⅜	4½	10⅜	11½
8	232	19½	27⅝	18	17½	17½	5½	4⅜	10⅝	11⅞
9	211	20⅛	28⅛	18	17⅞	17⅝	5½	4⅜	11	12⅛
10	209	20⅜	28½	18¼	18¼	17⅝	5⅝	4⅜	11¼	12⅜
11	193	21⅛	29	18⅜	18⅜	17⅞	5⅞	4⅜	11½	12⅝
12	284	21⅞	29⅜	18⅜	18½	17⅞	5⅞	4½	11⅝	13
13	264	22⅛	29⅝	18½	18⅝	18¼	5⅞	4½	11⅞	13¼
14	212	23	30¼	18½	18⅝	18⅜	5⅞	4½	12	13⅜
15	215	23⅛	30⅜	18½	18⅝	18½	6	4½	12¼	13⅜
16	186	24⅛	31⅛	18⅞	18⅞	18⅝	6	4⅝	12½	13⅝
17	159	24⅛	31⅜	18⅞	18⅞	18⅝	6	4⅝	12⅝	14
18	181	24⅝	31½	19⅛	19⅛	18⅝	6	4⅝	12⅝	14⅛
19	156	25⅛	32⅛	19⅛	19⅛	18⅝	6⅛	5	12⅞	14⅜
20	170	25⅜	32⅜	19⅛	19¼	18⅞	6¼	5	13⅛	14⅜
21	159	25⅜	32⅝	19¼	19¼	18⅞	6¼	5	13⅛	14⅝
22	160	26⅛	33⅛	19⅜	19⅜	18⅞	6¼	5	13⅜	15
23	184	27	33⅜	19⅜	19⅜	18⅞	6¼	5	13⅜	15⅛
24	201	27⅛	33¾	19⅜	19⅜	19	6¼	5⅛	13½	15⅜
25	150	27⅛	34	19½	19⅜	19⅛	6¼	5⅛	13½	15⅜
26	147	28¼	34⅛	19½	19⅝	19⅛	6¼	5⅛	13⅝	15⅝
27	126	29	34¼	19⅝	20	19¼	6⅜	5⅛	14⅛	15⅞
28	121	29⅛	35⅛	19⅝	20	19⅜	6⅜	5⅛	14⅛	16⅛
29	94	29⅛	35⅜	19⅝	20	19⅜	6⅜	5⅛	14⅛	16⅛
30	153	29⅜	35⅝	19⅝	20⅛	19⅜	6⅜	5⅛	14⅜	16⅜
31	113	30⅛	35⅞	19¾	20⅛	19⅜	6⅜	5⅛	14½	16⅜
32	110	30⅜	36	19¾	20⅛	19⅜	6⅜	5⅛	14⅜	16⅜
33	102	30⅜	36⅛	19¾	20⅛	19⅜	6⅜	5⅛	14⅜	16⅝
34	113	31⅛	36⅜	19⅞	20⅛	19⅜	6⅜	5⅛	14⅞	17⅛
35	116	31⅛	36⅝	19⅞	20⅜	19⅞	6⅜	5⅛	15⅛	17⅛
36	76	32⅛	37⅛	19⅞	20⅜	19⅞	6⅜	5⅛	15⅛	17⅜
37	25	32⅛	37⅜	19⅞	20⅜	20	6⅜	5⅛	15⅛	17⅜
38	37	32⅜	37⅝	19⅞	20⅜	20	6⅜	5⅛	15⅜	17⅜
39	36	33⅛	37⅞	20	20⅜	20	6⅜	5⅛	15⅜	17⅝
40	24	33⅜	38⅛	20	20⅜	20⅛	6⅜	5⅛	15⅜	18⅛
41	21	33⅜	38⅛	20	20⅜	20⅛	6⅜	5⅛	15⅜	18⅛
42	27	33⅜	38⅜	20	20⅜	20⅛	6⅜	5⅛	15⅜	18⅛
43	17	33⅜	38⅜	20	20⅜	20⅛	6⅜	5⅛	15⅜	18⅜
44	19	34¼	38⅝	20	21	20⅛	6¼	5⅛	15⅛	18⅜
45	27	34⅛	39	20	21	20⅜	6¼	5⅛	15⅛	18⅜
46	17	34⅛	39	20	21	20⅜	6¼	5⅛	16	18⅜
47	21	35⅛	39⅛	20⅛	21	20⅛	6¼	5⅛	16	18⅞
48	19	35⅛	39½	20⅛	21⅛	20⅜	6¼	5⅛	16	18¼

Total number of males measured, 5,602.

(a) All measurements are in inches except weight.

Digitized by Google

AVERAGE PHYSICAL MEASUREMENTS. (a)
FEMALES.

Age, Months.	No.	Weight, Pounds.	Height.	Circumference of Head.	Chest.	Abdomen.	Diameter of Chest. Lateral.	Ant.-Post.	Length of Arm.	Leg.
6	176	16¼	25½	16½	16½	16¾	5½	4½	9½	11½
7	188	17⅜	26½	17½	17½	17	5½	4½	10⅛	11⅜
8	186	18½	27	17⅝	17⅝	17½	5⅝	4⅝	10⅜	11½
9	183	19⅛	27⅛	17⅝	17⅝	17⅝	5⅝	4⅝	10⅝	12
10	175	19½	27⅞	17⅞	17⅝	17⅝	5⅝	4⅝	10⅞	12⅛
11	165	20⅛	28⅛	17⅞	17⅞	17⅞	5⅞	4⅝	11⅛	12⅜
12	228	20½	28½	18⅛	18	17⅞	5⅞	4⅝	11⅜	12⅝
13	208	21	29⅛	18⅛	18⅛	17⅞	5⅞	4⅝	11⅝	13
14	208	21⅛	29½	18⅛	18⅛	18	5⅞	4⅝	11⅝	13⅛
15	197	21⅜	30⅛	18⅛	18⅛	18	5⅞	4⅝	11⅞	13⅜
16	176	22⅛	30⅛	18⅛	18⅛	18	5⅞	4⅞	12	13⅜
17	143	22⅜	30⅜	18⅛	18⅛	18	5⅞	4⅞	12⅛	13⅝
18	178	23⅛	31⅛	18⅛	18⅛	18⅛	6	4⅞	12⅛	14
19	132	23⅜	31⅛	18⅛	18⅛	18⅛	6	4⅞	12⅜	14⅛
20	131	24⅛	32	18⅛	18⅛	18⅛	6	4⅞	12⅝	14⅛
21	142	24⅛	32⅛	18⅛	19	18⅛	6⅛	4⅞	12⅞	14⅜
22	125	25⅛	32⅜	18⅜	19⅛	18⅛	6⅛	4⅞	13	14⅜
23	134	25⅛	32⅜	19	19⅛	18⅛	6⅛	4⅞	13⅛	15
24	160	26⅛	33⅜	19	19⅛	19	6⅛	4⅞	13⅜	15⅛
25	126	26⅛	33⅛	19	19⅛	19	6⅛	4⅞	13⅜	15⅛
26	131	27⅛	33⅜	19⅛	19⅛	19⅛	6⅛	5	13⅜	15⅛
27	136	27⅛	33⅜	19⅛	19⅛	19⅛	6⅛	5	13⅜	15⅛
28	95	27⅛	34⅛	19⅛	19⅛	19⅛	6⅛	5	14	16
29	95	27⅛	34⅛	19⅛	19⅛	19⅛	6⅛	5	14	16⅛
30	132	28⅛	34⅛	19⅛	19⅛	19⅛	6⅛	5	14	16⅛
31	109	28⅛	35⅛	19⅛	19⅛	19⅛	6⅛	5	14⅛	16⅛
32	116	29	35⅛	19⅛	19⅛	19⅛	6⅛	5	14⅛	16⅛
33	97	29⅛	35⅛	19⅛	19⅛	19⅛	6⅛	5	14⅛	16⅛
34	91	30⅛	36⅛	19⅛	20	19⅛	6⅛	5	14⅛	16⅛
35	84	30⅛	36⅛	19⅛	20	19⅛	6⅛	5	14⅛	16⅛
36	90	30⅛	36⅛	19⅛	20	19⅛	6⅛	5	14⅛	17⅛
37	28	30⅛	36⅛	19⅛	20	19⅛	6⅛	5	14⅛	17⅛
38	31	31	37	19⅛	20⅛	19⅛	6⅛	5	14⅛	17⅛
39	37	31⅛	37⅛	19⅛	20⅛	19⅛	6⅛	5⅛	15	17⅛
40	26	32	37⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	17⅛
41	16	32⅛	37⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	18
42	42	32⅛	38	19⅛	20⅛	20	6⅛	5⅛	15⅛	18⅛
43	12	32⅛	38⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	18⅛
44	22	33	38⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	18⅛
45	18	33⅛	38⅛	19⅛	20⅛	20	6⅛	5⅛	15⅛	18⅛
46	9	33⅛	38⅛	19⅛	20⅛	20⅛	6⅛	5⅛	15⅛	18⅛
47	12	33⅛	38⅛	19⅛	20⅛	20⅛	6⅛	5⅛	15⅛	18⅛
48	31	33⅛	39	19⅛	20⅛	20⅛	6⅛	5⅛	15⅛	18⅛

Total number of females measured, 4,821.

(a) All measurements are in inches except weight.

age than girls. It is interesting also to note the relative changes in the circumferences of the chest and abdomen. The circumference of the abdomen in very young children generally exceeds the circumference of the chest, and this is frequently true at ages four to six months, but with increasing age up to nine months there is only a slight difference in the measurements. This is true of both sexes. From the age one year upward the excess of circumference of the chest over the circumference of the abdomen is greater for boys than for girls. For illustration, at age 24 months the average chest circumference of boys is 19⅛ inches, and the average abdomen circumference of boys is 19 inches. For girls of the same age, the chest circumference is 19⅛ inches, and the abdomen circumference is 19 inches.

Digitized by Google

The head circumference of boys was found to average approximately ½ inch more than the head circumference of girls, and this difference is persistent throughout the period of life covered by the table.

From this table the average rate of growth can be determined with approximate accuracy for any duration within the limits of the table, and the comparative rate of growth of different parts of the body is likewise ascertainable for both sexes. The great majority of the children whose measurements are included in this table were of American-born parents; the children, however, were of different stocks, including German, Irish, Swedish, some Italian, and some of various other races. It would be desirable to have similar tables for the different racial elements, and perhaps even for different sections of the country. There is some evidence that the average measurements, even of these young children, would be found to differ, in the South, for instance, as compared with the North and West, entirely aside from the probable racial differences. It would seem, however, that this table, being based upon measurements of normal healthy children in various sections of the country, should serve as a fair guide for many practical purposes without any additional refinements.

FREDERICK S. CRUM.

A PLAN FOR GATHERING STATISTICAL DATA AS A BY-PRODUCT OF ADMINISTRATIVE WORK.

We, in this country, have for a number of years been familiar with the using for scientific purposes of records kept for administrative purposes. Thus the records kept by charity organization societies, chiefly in order to increase the efficiency of their work with individual families, have been extensively used as data for the scientific study of the causes of poverty and dependency.* In some cases also items or questions, which were not necessary for administrative purposes, have been added to the record forms with the purpose in view of using the data as a basis for scientific studies of cause and effect. In such cases the supplementary data recorded for scientific uses are a sort of by-product, turned out in the regular course of administrative record keeping.

This plan of gathering data as a by-product to regular administrative work has not been so extensively used in field investigations. In at least one instance, however, effective use has been made of it—by Dr. John Robertson, Medical Officer of Health for Birmingham, England, in his study of the relation of the employment of mothers to infant mortality. The advantage of his method appears to be sufficiently great to warrant a brief description of it.

* See for example Devine's "Misery and Its Causes," pp. 230 ff., and Conyngton's "How to Help," pp. 15-16, 342-3.

Digitized by Google

In describing the plan used Doctor Robertson said in his report for 1909: "On May 10, 1907, a letter was received from the Home Office to the effect that the Home Secretary had under consideration the question of the further regulation of the industrial employment of women before and after child birth. . . . In order that fuller information as regards the effect of employment both before and after child birth on the health of the mother and child . . . might be obtained, a scheme of investigation was drafted and medical officers of health were asked to take part in the collection of data. . . . ".

Having decided that two wards in his district (St. Stephen's and St. George's), because of the large number of women gainfully employed, furnished an excellent area for such an investigation, a plan or method of collecting the data had to be devised. The method chosen was rather unique, although Doctor Robertson does not appear to have appreciated this fact. Instead of employing investigators to make special visits to the homes of mothers for the particular purpose of obtaining the information needed, it was decided to simply have the health visitors, who in the course of their regular work would be visiting the mothers in the two wards anyway, ask the necessary questions and make the needed observations as a supplement to their regular duties. Thus the gathering of the data as to the relation of the employment of mothers to infant mortality was a sort of by-product of the regular health work.

The visits were made by Dr. Jesse Duncan and "two experienced health visitors." "Every baby born (with a few exceptions) was visited at "frequent intervals" and at the end of twelve months weighed. Describing the results for 1908 Doctor Duncan says: "The children born in the district are visited as soon as convenient. . . . At this visit directions are given [as usual] as to the care of the child, but [in addition] information is obtained regarding the mother's employment, previous history, husband's wages," the mother's age, the method of feeding, etc. This plan of investigation was also followed in 1909 and 1910.*

From the standpoint of accuracy and thoroughness, the advantages of this method of gathering data are clear. Doctor Robertson shows this when he says: "I desire to add that the work which Dr. Jesse Duncan [assisted by the two health visitors] has done, and upon which this report is based is characterized by its accuracy and thoroughness. She has spent her whole time among those residing in St. Stephen's and St. George's Wards, and is familiar with every phase of life in these districts. This has enabled her to check and correct the statements made by the mothers she has interviewed." This is one advantage. In another place he points out another—the fact that "close contact was maintained with each of these mothers during a whole year, and much information not available at the first visit was obtainable at subsequent visits."

* Health Department of the City of Birmingham, England: "Report on Industrial Employment of Married Women and Infant Mortality," supplement to the annual report of the department for 1909, and "Report on Infant Mortality in St. Stephen's and St. George's Wards," supplement to the annual reports of the department for 1910 and 1911.

Digitized by Google

The expense involved in gathering data by this method is probably also considerably less than that involved in the usual house to house method. Since the data gathered is a sort of by-product of the regular work of the health visitors and since the visits to the homes would be made and probably most of the necessary questions asked any way, even if no records were kept of the answers, it would appear that the additional expense involved would be chiefly that of recording and tabulating the data and preparing and printing the report.

This method of gathering data might well be adopted by many American cities in studying infant mortality and similar problems.

H. H. Hibbs, Jr.

Nashville, Tenn.

India. Moral and Material Progress and Condition of India During the Year 1913–1914. India Office. London: Eyre and Spottiswoode, July, 1915. Pp. 137.

This report is compiled from detailed accounts from each of the presidencies and districts and is presented to parliament annually. The statement is divided into 18 chapters, which summarize administrative and financial matters, sources of revenue, condition of various industries, and other aspects of life in India.

The remarkable growth of coöperative credit societies is outlined. These were primarily organized to help the cultivators and to protect the poorer classes from greedy and merciless money lenders. The membership increased from 28,629 in 1905–6 to 744,226 in 1913–14, and the working capital from £32,000 to £5,145,000. These societies help to develop not only rural communities, but urban as well. The societies, through small loans, enable artisans to purchase necessary tools and implements. This fact is of great importance since the manufacturing population has increased sufficiently to necessitate special labor legislation. In a good many respects this legislation is not much, if at all, behind that of some of our states. Factory inspection and condition of women workers are definitely covered. Child labor laws call for working papers (issued upon medical examination); a minimum age for employment (9 years); and maximum number of working hours (6 hours in textile industry, 7 in others). These facts speak not only of industrial development, but of intellectual and social growth as well. This is further emphasized by the chapter on education which shows that colleges and institutions of even higher learning can boast not only of native male, but also female students.

Defects in the vital statistics are frankly admitted. In addition to the usual causes of under registration met with in other countries,—indifference to and neglect of legislation—the Government in India is confronted with utter ignorance on the part of information gathering agencies as well as on the part of the population as a whole. Even with defective registration, the records indicate a very high birth rate (39.4 per 1,000). It is due,

Digitized by Google

however, not so much to fertility as to a large proportion of married women in the population. The death rate is also very high (28.7 per 1,000). The nature of the chief causes of death indicates that the high rate is due principally to unsanitary habits and conditions. The following table presents the principal facts of deaths by diseases:

MORTALITY FROM PRINCIPAL CAUSES OF DEATH IN INDIA, 1911-1913.

Diseases.	1913.		1912.		1911.	
	Number of Deaths.	Rate per 1,000.	Number of Deaths.	Rate per 1,000.	Number of Deaths.	Rate per 1,000.
Fevers (a)	3,983,112	16.71	3,936,085	16.49	4,207,356	17.63
Cholera	294,815	1.24	407,769	1.71	354,005	1.48
Dysentery and Diarrhoea	246,578	1.03	292,216	1.22	253,636	1.06
Respiratory Diseases	237,229	1.00	247,736	1.04	223,822	.94
Plague	198,456	.83	263,037	1.10	733,582	3.07
Smallpox	98,155	.41	89,357	.37	59,538	.25

(a) Malarial fever mostly.

The existing state of affairs, bad as it is, indicates a great improvement, thanks to the growing number of hospitals, dispensaries, Pasteur institutes, and other public welfare institutions.

Another vital factor in the life of the people is the constant internal migration of population. It is of such proportions that it affects the scale of wages which, according to the report, keeps a fairly even pace with the rise of prices. Wages are further influenced by an increasing demand for labor due to industrial development.

The report shows definite progress,—steady if slow. The Government office is to be commended for its care and perseverance in gathering the facts under conditions which are most unfavorable for obtaining reliable information, both through limited facilities and ignorance of informers.

JOSEPH RAYWID.

Switzerland-Zeitschrift fur Schweizerische Statistik, 50. u. 51 Jahrgang. Bern: Stampfli and Company. 1914 and 1915.

The fiftieth volume of this journal of Swiss statistics contains Dr. Karl Bader's article on "Groups in Statistics and in Sociology," a philosophy of classification. Heinrich Schlosser of the Museum of Industrial Arts, Zurich, discusses the application of the elements of design to the construction of graphic charts. In a review of graphic statistics at the Swiss Exposition, Bern, 1914, Dr. Hans Schorer defines his ideas of graphic illustration of statistical data. Both of these articles should interest American statisticians because of current activities in the standardization of graphic statistical practice on the part of American learned societies.

The fifty-first volume gives a review of official statistics at the 1914 Swiss Exposition by Dr. C. Muhlemann. The same article discusses the

Digitized by Google

nature and purposes of official statistics as adjuncts to government, legislation and the political sciences in general. Dr. Furlan of Basel treats the subject of reorganization of the Swiss statistical service in his article and recommends the formation of a Central Statistical Commission for the republic.

Vital statistics for the year 1914 are also given. The European War profoundly affected the marriage rate; the effect upon birth and death rates will not become statistically apparent until the data for 1915 are published.

An article published by the Swiss Statistical Bureau shows the results of the 1910 population census for the several main classes of the population with respect to country of origin, and sex and age.

The effect of the war upon Swiss industries may be seen in the statistics of manufactures and of exports in the tabular appendix to the 1915 Journal. Furlan, in another study, presents his investigation of the density of the Swiss population.

E. W. KOPF.

New York City.

1920 CENSUS IN ITS RELATION TO VITAL STATISTICS.

Dr. John W. Trask, Chairman of the Vital Statistics Section, American Public Health Association, has appointed the following committee to take under consideration the subject of vital statistics of the 1920 United States Census: Dr. W. H. Guilfoy, Chairman, Dr. Cressy L. Wilbur, Dr. John S. Fulton, Prof. Jas. W. Glover, Mr. Edwin W. Kopf, Prof. Walter F. Willcox, Mr. Frederick L. Hoffman, Mr. Franz Schneider, Jr., Dr. W. H. Frost, and Dr. Wm. H. Davis.

Among the problems which the committee will be called upon to consider carefully are:

Life Tables for the triennium, 1919–1921.

Differential fecundity and vitality of race and nativity classes of the population.

Revision of mortality statistics of the decade 1910–1919 in the light of 1920 Census results, including review of achievements in reducing mortality from preventable disease.

Mortality in occupations, triennium 1919–1921.

Service of vital statistics in social reform; sickness, mortality and other vital statistics as a foundation for social insurance.

Vital statistics of institutions for care of the sick, triennium, 1919–1921. (Hospitals, asylums for the insane, indigent, etc.)

Plans for the improvement and extension of the registration area for births and deaths.

Improvement of municipal and state vital statistics.

Revision of International Nomenclature and Classification of Diseases, applicable to the years 1919 *et seq.*

Digitized by Google

Enumeration of sicknesses on Census Day, 1920, in selected areas.

Sickness and accident statistics of insurance societies.

The necessity of the census authorities furnishing some of the large cities of the country with age groupings of the population according to nativities and other related facts.

The age groupings according to selected areas in these cities.

The age groupings according to occupations.

The appointment of the committee recalls the symposium on vital statistics for the Twelfth Census, 1900, which appeared in the March, 1897, Quarterly of this Association and contributed by Messrs. Wilbur, Willcox, Fouse, Hoffman, Abbott, and King.

E. W. KOPF.

New York City.

Digitized by Google

AMERICAN PUBLIC HEALTH ASSOCIATION.

PRELIMINARY PROGRAM OF THE SECTION ON VITAL STATISTICS, OCTOBER 24–27.

Chairman, DR. J. W. TRASK,
Washington, D. C.

Secretary, DR. LOUIS I. DUBLIN,
New York, N. Y.

FIRST SESSION.

Symposium on Enforcement of Registration of Births and Deaths and the Use of the Resulting Records.

1. The Enforcement of Registration of Births and Deaths in Metropolitan Areas—Methods, Results and Checks. Wm. H. Guilfoy.
2. The Enforcement of Registration of Births and Deaths by a State Registration Office—Methods, Results and Checks. Cressy L. Wilbur.
3. The Use of Registration Records of Births and Deaths in the Commercial, Scientific, and Legal Life of Metropolitan Populations. S. W. Wynne.
4. The Use of Mortality and Morbidity Statistics in Epidemiologic Studies. Allan W. Freeman.
5. The New United States Life Tables—Their Purpose and What They Show. Jas. W. Glover.
6. The Enforcement of Registration of Births in Indiana. J. N. Hurty.

SECOND SESSION.

Symposium on the Notification of Disease and the Utilization by the Health Department of the Information thus Obtained.

1. The Uses Made of Morbidity Reports by the Minnesota State Department of Health. A. J. Chesley.
2. The Uses Made of Morbidity Reports by the New York State Department of Health. F. M. Meader.
3. The Notification of Disease and the Uses Made of the Information thus Obtained in Municipal Public Health Administration. W. H. Price.
4. The Notification of Tuberculosis and the Supervision of Reported Cases in Massachusetts. Eugene R. Kelley.
5. The Establishment of "An Area of Known Disease Prevalence." Is it practicable and desirable at the present time? What diseases should be reported and what standards of enforcement of notification required? John S. Fulton.
6. The Registration of Mental Disorders. A. J. Rosanoff.

Digitized by Google

THIRD SESSION.

(Joint Session with Sociological Section.)

Symposium and Discussion on Methods of Collection and Expression of Data in Health and Social Investigations.

1. The Necessity for Health Standards. Dr. Donald B. Armstrong, Director, Department of Social Welfare, New York Association for Improving the Condition of the Poor.
2. The Limitation and Necessities of Socio-Health Investigations. Mr. Frans Schneider, Sanitarian, Department of Surveys and Exhibits, Russell Sage Foundation, New York City.
3. The Application of Statistical Methods to the Field of Socio-Health Research. Dr. Louis I. Dublin, Statistician, Metropolitan Life Insurance Co., New York.
4. Discussion: Led by J. W. Schereschewsky, Surgeon, United States Public Health Service; Dr. Ernest C. Meyer, International Health Commission, New York City; Dr. Haven Emerson, Commissioner of Health, New York City; Prof. C.-E. A. Winslow, Yale University, New Haven, Conn.; Prof. E. B. Phelps, Hygienic Laboratory, Washington, D. C.; Dr. W. C. Woodward, Commissioner of Health, Washington, D. C. Other discussants to be announced later.

FOURTH SESSION.

1. Reports of Committees.
2. Vital Statistics and Publicity as Exemplified by New York City's Recent Experience with Poliomyelitis. Chas. F. Bolduan.
3. The Massachusetts Endemic Index. H. R. Brown.
4. The Seasonal Distribution of Communicable Diseases Based on Morbidity Experience. John S. Fulton.
5. A Statistical Study of Diphtheria. F. S. Crum.

All communications with regard to this section should be addressed to Dr. Louis I. Dublin, 1 Madison Avenue, New York City.

Digitized by Google

Digitized by Google

AMERICAN
STATISTICAL ASSOCIATION

NEW SERIES, No. 116. DECEMBER, 1916.

A STUDY IN SOCIAL DYNAMICS: A STATISTICAL DETERMINATION OF THE RATE OF NATURAL INCREASE, AND OF THE FACTORS ACCOUNTING FOR THE INCREASE OF POPULATION IN THE UNITED STATES.

By John M. Gillette, Ph.D., *Professor of Sociology, University of North Dakota.*

Little appears to have been published relative to the specific points discussed in this study. The present writer has become interested in the field in consequence of the repeated attempts he has made to ascertain the amount of rural migration. In seeking to devise methods by which that factor could be measured, he has gradually developed conceptions as to how the larger problems of natural increase for the nation and its parts and how the factors which determine any particular kind of community increase may be measured. His earlier publications of estimates of rural migration unfortunately were inaccurate, but a later study* in this present field, which embraced a calculation of that factor also, laid a more ample and trustworthy foundation for the present investigation. In the same issue Mr. Earl Clark, of the Russell Sage Foundation, published an estimate of the factors of urban increase in the United States,† the results of which, when account is taken of the somewhat different view of the scope of the undertaking, practically agree with those embodied in the study just cited.

Investigations relative to natural increase of population in

*"Measure of Rural Migration and Other Factors of Urban Increase in the United States," by John M. Gillette and George R. Davies, Quarterly Publications of the American Statistical Association, Vol. XIV, pp. 640-653.

†"Contributions to Urban Growth," by Earl Clark, *Ibid*, pp. 654-671.

Digitized by Google

the nation as a whole, in its several geographical divisions, and in its distinct types of communities, rural and urban, which further attempt to analyze the situation relative to those different areas and communities so as to exhibit the comparative strength of the various determining and contributing factors affecting the increase or decrease, are of the profoundest significance. They are far beyond being of "academic" interest only, for they reveal the mechanism of the national population at work and the basis for a causal interpretation in the field of economics, political science, sociology, history, ethics, and other related sciences. They throw light on and go far towards answering the question as to what stocks in our population, native or foreign born, are the vigorously reproductive ones. They yield data which will do much to determine whether or not immigration to this country has actually increased the population beyond what it would have been had the nation depended on its own reproductive powers alone. They also help to answer the query as to which kind of community, rural or urban, is the more advantageous with respect to the increase of population. In like manner they reflect the determinative influence which economic opportunities have on natural increase as exhibited in the varying rates of the different geographical divisions. Further, these results reveal unmistakable differences between native and foreign stocks, and the Negro and white races, in their competition for survival.

Such considerations and results invest the data with a practical concern of the utmost gravity. They project into the governmental sphere such questions as these: what stock, regions, and kind of community are most advantageous when measured by their effects on the fecundity of population? Since the results are what they are, what should the governments, national, state, and municipal, undertake to ward off the decadent tendencies and to place a premium on the beneficial ones?

They also project the same practical questions into the arena of individual consideration, and arouse serious personal concern. In view of the unfavorable conditions that must be inferred to exist in some regions or in some kinds of community,

Digitized by Google

should he seek to cast his lot in the region or community which is proven to be the more favorable to human life? The vast interdivisional migration of native born populations indicates that already millions of people are instinct with this consideration. Should other millions once become intelligent and acquainted with the actual conditions, a movement might result that would change the balance of population and affect the destiny of great areas.

The difficulties involved in making estimates in this field are great. Because of the almost overwhelming complexity of the situation, the multitudes of unknown factors involved, and the dearth of exact and pertinent data, the project is not dissimilar to that of the celestial physicist who, not being able to observe the disturbing body, is obliged to locate it by a long series of mathematical calculations.

The veteran statistician knows perfectly well that an investigation of this nature develops problem after problem for which there exists no ready made solution and which the ingenuity of the investigator must meet by devices and methods wrought out as occasion demands. The search for facts which will give a cue to the solution of such problems is time-consuming and frequently disappointing. Circuitous methods are often necessary which many times lead the zealous amateur worker into blind alleys. It may be glorious to spend a month of earnest effort to work out results by a newly invented device, but if they are worthless it is indeed discouraging. There is no glory unless success is at last attained. The writer has nearly 200 pages of legal notepaper filled with the mere indication of methods and results developed in the course of this investigation, and the mathematical operations to carry out the methods would demand a far greater space. It has often been a case of trial and error. Perhaps the reason the task has not been executed earlier was that the veteran statistician suspected the risks involved in the project and the foolhardiness of the undertaking.

The statistician does not expect absolute accuracy in a set of statistical results because he knows that statistics is the science of estimates and that the final figures are only approximations. The only claim that a statistical in-

Digitized by Google

vestigator dare make is that his conclusions come close to the true situation.

The writer has sought to work out scientific methods by means of which the problems set up might be solved. That there is such close conformity in several sets of results which were worked out by utterly different methods is the most convincing evidence to his mind that the conclusions are approximately correct. This is particularly true respecting the rates of natural increase attained for the United States. Those worked out by means of data and methods quite independent of each other show a divergence of only about five one hundredths of one per cent. And much the same is true of the division rates independently obtained, and of the totals of the factor, rural migration, resulting from several distinct operations.

I made certain modifications in mortality rates, as a concession to Professor Walter F. Willcox, after I read his strictures on low mortality rates reported by some of our western states.* I revised my estimates in several states for both rural and urban districts, raising them somewhat above my first estimate. Since the rates for these communities, for the different divisions, and for the nation are the weighted averages of the local rates, the final rate of natural increase for the nation was only slightly lowered.

No doubt the general scheme adopted here will undergo a modification in certain of its particulars in the hands of critical investigators. But the present undertaking should serve as a foundation and point of departure for future workers in the attainment of similar results. When the next census data are made available, it will be possible to apply the plan developed here to these data and to secure results with a minimum cost in time and labor.

By natural increase is meant, for the purposes of this paper, the increase that was made during the decade 1900–1910 by the population of 1900, independent of the gains from immigration. Since my chief aim is to reckon the potency of our various population areas and communities, the factor of emigration must be considered, since to disregard it would be to

* Quarterly Publications of the American Statistical Association, Vol. XIV, pp. 711–726.

Digitized by Google

distort the idea of our real producing ability. This conception of natural increase differs from the one commonly employed, since the latter includes all actual gains by natural increase to the initial population, whether due to its own reproductive processes or to those of immigration.

A considerable degree of positivity attaches to the rate of natural increase estimated for the nation because that rate has been established by two different methods. The first method consists in the use of immigration and emigration to determine what part of the decennial growth of population was due to natural increase and what was the potency of the latter. The second method seeks to establish birth and death rates for the nation, the difference between the birth rate and the death rate being the rate of natural increase. It will be seen that two different estimates of the birth rate are employed and that these are variants of the second method.

The national population in 1900 was 75,994,575; in 1910 it was 91,972,266, a gain of 15,977,691.* This increase of 21.0 per cent. can not be the rate of natural increase of the nation because of the disturbing factors of immigration and emigration.

The factor of emigration necessarily enters into the process of estimating the rate of natural increase of the nation, when the determination of that increase involves the computation of the relative force of the various causes which produce gains and losses in the population. Emigration counts as a loss factor, but nevertheless it is a positive element in calculating the potency of natural increase.

Strange as it may seem our national government has done practically nothing toward recording losses of our permanent population. It keeps a record of emigrant aliens, but statistics which could form a basis of judgment as to the amount of emigration of citizens are scanty and only within the last few years have these data appeared. Consequently my first attempt at estimating the force of this factor was based on the statements of emigration from the United States to Canada published by the Canadian government. The results of the estimate appeared in a recent article in the Publications of the

*Abstract of 13th Census, p. 22.

Digitized by Google

American Statistical Association.* It was reckoned that during the decade 1900–1910, 517,000 individuals settled in Canada† and that enough more went to other parts of the world to bring the total emigration up to 550,000. The present calculations indicate that this amount is approximately correct.

The more recent reports of the United States Commissioner General of Immigration‡ afford data from which an approximation can be made. They record the number of citizens of the United States annually arriving and departing at ports of entry and at the Canadian boundary stations. During the four years, 1911–1914, the excess of departures over arrivals was 296,742. For a ten year period at the same rate, the excess would total 741,830. Assuming that the same relative emigration occurred during the preceding decade, and proportioning the excess to the average population of that decade, the resulting emigration for 1900–1910 is found to be 557,000. This so nearly corresponds to my previous estimate that that estimate may be used unchanged.

We may anticipate the rate of natural increase in the nation to apply to this amount of emigration for finding its total reducing power on our population. Had the 550,000 persons remained in the United States their number would have been increased by one half of 13.7 per cent. making the total loss 588,000.

Since 1820 immigration has been an increasing factor in the growth of our population. The United States has exhibited the most rapid rise in population of all the modern nations, Russia standing as a close second. It is a matter of great importance to determine to what degree our national population of itself contributes to this remarkable advance. How far will immigration account for the decennial rate of gain of 21 per cent.?

In estimating the contribution of immigration, since I am determining the potency of the 1900 population, it will be necessary to reckon the decennial net immigration, and the

*Sept., 1915, p. 645.

†"Immigration Facts and Figures," issued by the Minister of the Interior, Ottawa, Canada, p. 3, gives data for the Canadian estimate.

‡Reports of the Commissioner General of Immigration, United States, 1912, p. 67, and 1914, p. 35.

Digitized by Google

natural increase for that immigrant population. The first task is simple, the other one difficult because of great complications.

The United States Bureau of Census estimates the net immigration of the decade, 1900–1910, as 5,250,000.* This closely agrees with the results obtained by reckoning that 80 per cent. of the foreign born within the United States in 1900 lived in the United States in 1910† and subtracting the amount from the number of foreign born enumerated in 1910. This gives a net immigration of approximately 5,243,000.

It is necessary to discover how much this net immigration is incremented by its own natural increase during the decade in question. My estimate of 17.4 per cent. as the rate of natural increase for that population was obtained as follows:

The birth rate for immigrants was reckoned by taking the European birth rates of Hungary, Prussia, Germany, Italy, Great Britain, Norway, Sweden and Denmark, France, Spain, Portugal, and Ireland, nations from which we obtain the bulk of our foreign population, and weighting them by the appropriate populations living in the United States. This was found to be 32.8 in 1900.‡ Had we the figures for Russia, the average rate doubtless would be greater.

This average rate of 32.8 must be increased because of the excessive adultness of immigrants. As compared with the population of the United States in 1910, the immigrants of the decade ending 1910 were 59 per cent. more adult, that is, were of the fecund age, 15 to 45, to an extent 59 per cent. greater than the whole population.§ So increased, the birth rate becomes 52.2.

But immigrants have a preponderating maleness. I estimate that females constitute 38.6 per cent. of the net immigrant population.‖ Assuming a male for each female, there is an excess of 22.8 per cent. of males over females. The last resulting rate, reduced by this percentage, is 40.5. It is pos-

*Abstract 13th Census, p. 191.

†Census 1910, Polulation, Vol. I, p. 1017.

‡Abstract 13th Census, p. 188, and Bailey, Modern Social Conditions, p. 97.

§Abstract 13th Census, p. 128.

‖Estimate from Reports Commission General of Immigration, United States, 1909, 1910, 1911, 1912, 1913.

Digitized by Google

sible that this rate is too low.* On the other hand, since the birth rate of populations migrating to communities having higher standards of living tends to decrease and since most of the immigrants are derived from rural Europe and settle in cities, it may be concluded that greater economic opportunities here do not largely increase the rate.

A further estimate is necessary before this rate can be applied to the immigrant population. The net immigration of 5,243,000 suffered a death rate during the decade, the decedents being eligible to increasing the population before death. On the basis of the average mortality rate of that decade there should have been enough decedents to swell the child bearing immigration to 5,651,000. From the application of the above birth rate, it is found there would be born 1,130,000 children during that period of time.

But the children born to immigrants suffer a decrease by death during the ten year period. According to a probability curve drawn from Forsythe's Life Tables,† about 18 per cent. of the population dies by the tenth year of age. While this rate may be too low for the class of children in question it must be used because of the dearth of other figures. This reduces the number of children to 916,500. Dividing this sum by 5,243,000, the decennial increase rate for the immigrant population of 17.4 per cent. is obtained. As the result of this series of computations concerning immigration, it is found that there is a numerical population value of the specific net immigration amounting to 6,159,000.

The entire potency of the 1900 population for increase is measured by decreasing the decennial gain in population, 15,977,691, by the total force of the net immigration, 6,159,000, and increasing the result in turn by the amount of emigration plus its increase, 588,000. Dividing the resulting 10,407,000 by the 1900 population, 75,994,575, the resulting percentage, 13.7, is the rate expressing the power of that population to reproduce.

The actual increase, however, is expressed by the percentage obtained by disregarding emigration and the natural increase

*See Ellwood, Sociology and Modern Social Problems, p. 176, and Report of Immigration Commission, Vol. 2, p. 497.

†Quarterly Publications, Am. Stat. Assn., Sept., 1914, p. 234.

Digitized by Google

of immigration, a percentage of almost 14.2. This latter percentage is slightly below the figure, 14.8, suggested by the census bureau at the 13th census.* And the potency rate of 13.7 per cent. is also below the census approximation of 14.0.

It is seen that immigration contributes a large share of the decennial increase of population. On the basis of the foregoing estimate its contributing power is $21-13.7\div21=34.7$, practically 35 per cent. During the three decades, 1790–1820, previous to the coming of immigrants in force, the decennial increase of population averaged 34.5 per cent., practically all of which was due to natural increase. That rate, less the present rate of 13.7, marks the decline in the rate of natural increase, a decline of over 60 per cent. of the original rate. The earlier rate is over two and a half times as great as the present rate. The consideration of the decline in the rate of natural increase yields a more direct and vivid picture of the movement of a modern nation in the direction of the much discussed "race suicide" than does a study of the decline in the birth rate, because, unless the latter is accompanied by similar study of the death rate, only hazy conclusions can be drawn. For real race suicide occurs only when the death rate exceeds the birth rate.

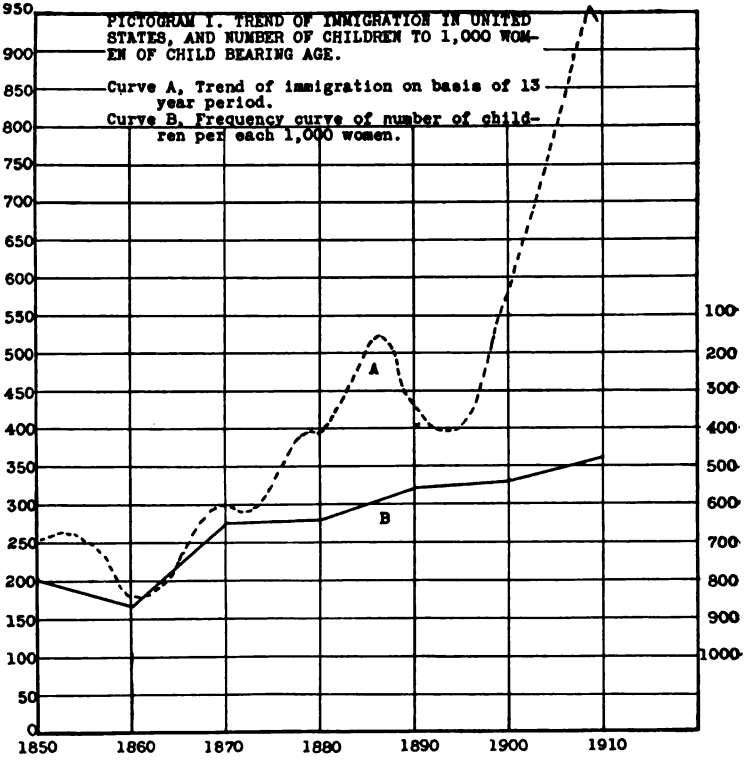
The question, as to whether or not immigration to this nation has really increased our population, might appear ridiculous in the face of the fact that during the last census decade that influx contributed 35 per cent. of the increase of our population, and likely accounted for quite as large a proportion of the gains during the more recent preceding decades. Yet some intelligent students of the question have so held, claiming that had not the immigrants come in competitive numbers our national rate of natural increase would have continued sufficiently high to offset our gains in population from immigration.

The fact is, however, that the birth rate of the United States was decreasing before the heavier waves of immigration struck our population. This heavy pressure evidently was not serious prior to about 1850. The ratio of foreign born to the population at that date was 1 foreign born person to 9.5 of

*Abstract 13th Census, p. 78.

Digitized by Google

the population in general. The next decade it rose to 1:7.6, from which it has not greatly varied since that time. The ratio was 1:19 in 1840 and 1:48 in 1830. The two latter ratios are estimated on the basis of the ratios of the successive decennial amounts of foreign born population from 1850 to 1870. Our birth rate doubtless would have decreased had there been only slight or no immigration, since other nations which have received little immigration have suffered a decided decrease of such rate. But it is somewhat apparent that immigration has affected the increase rate of the native stock. This is seen in the relation of the frequency curve (Graph I) of the number of children under 5 years of age per 1,000 women of 15 to 44 years of age, from 1850 to 1910, to the curve of the trend of annual immigration, drawn on the basis of a thirteen year



Digitized by Google

period, for the same time.* The variations of the former synchronize with the undulations of the latter in a striking manner. Of course, it is possible to say that the conditions which cause the fluctuations in immigration also produce similar changes in the rate of natural increase. No doubt there is a degree of truth in this statement.

Some have sought to demonstrate that it is urbanization which lowers the rate of population increase rather than immigration. Thus Goldenweiser offers selected statistics in support of this contention.† He places in parallel columns the "percentage of population living in the cities of at least 25,000 inhabitants: 1900," the "percentage of population foreign born: 1900," and the "number of children under 5 years of age per 1,000 native white women 15 to 44 years of age: 1900." Upon inspection, he pronounces an apparent correlation between the number of children and urbanization rather than between the former and immigration.

The correlation he considers apparent is not conclusive, however, for upon application of the scientific principle of correlation to his data it is discovered that the coefficient of correlation in favor of the influence of immigration on the number of children is greater than is that of urbanization, the former being .73 and the latter .66.‡ While, from these coefficients, a decisive case cannot be made out for either immigration or urbanization exclusively, yet since the coefficient for immigration is larger than that for urbanization it would appear that the former possesses the greater causal relation to the birth rate.

However, should we regard the factor of foreign born population as a more accurate measure of the pressure exerted on our native born population than the element of immigration, the above conclusion drawn from Goldenweiser's data would have to be reversed, since the coefficient of correlation between the foreign born factor and the number of children to 1,000

*For estimates of number of children, see article by Walter F. Willcox, QUARTERLY PUBLICATIONS of the American Statistical Association, XII, p. 491; the trend of immigration was made from data given in the Reports of the Commissioner General of Immigration.

†American Journal of Sociology, Vol. XVIII, pp. 348–350.

‡I desire to acknowledge my indebtedness to my colleague, Dr. George R. Davies, for suggesting the correlation between the frequency curve of children of fecund women and the fluctuations in immigration, and for the coefficients of correlation between the various factors in Goldenweiser's table.

Digitized by Google

women of child bearing age is only .38. Probably there is good reason for thinking of the pressure exerted by the foreign born on our native population as the truer index, since it is the permanent factor in the situation, immigration fluctuating according to conditions.

No doubt immigration has been a factor in causing a reduction of the birth rate in the United States, but that it is the sole factor cannot be sustained in the face of the fact that other states and nations having little or practically no immigration have sustained quite as large or greater decrease in their birth rates. That urbanization is an important causal factor in reducing the birth rate must be concluded from an inspection of the decennial rates of natural increase presented in columns 8 and 9 of Table I. When it is noted that the rural rate is almost twice the urban rate for the nation as a whole, that in only one division does the latter exceed the former, and that in some divisions the rural rate is three times the urban rate, it can scarcely be doubted that the factor of urbanization is the most important cause of lowered increase rates. Urban birth rates are lower than rural birth rates and its death rates are higher than those of the latter.

The task of securing rates of natural increase for rural and urban communities is a difficult undertaking, because data for their immediate determination are deficient. Had we mortality and birth data for them it would be a simple matter to obtain their difference. But trustworthy registration of births obtains for only New England and a few cities in other divisions, and the registration area for deaths embraces only about two thirds of the national population, many states and some divisions being without complete registration. And while nothing like an exact estimate can be made in view of these conditions, I believe it is worth while to make an initial calculation, trusting to time either to perfect it or to develop complete registration of births and deaths.

One method of securing natural increase rates for rural and urban districts is to find a substitute method of deriving births and to approximate death rates in states not having them.

If the census reported the number of children living in urban and rural communities under one year of age, we need only

Digitized by Google

to discover the portion of children who had died to obtain the number born. But the best the census does is to give the number of children under five years of age living in such districts.

By the use of a frequency curve constructed from data furnished by Forsyth's Life Tables* it is possible to calculate the average age and conversely the average number of children who died within the five year age group. The life table represents the elimination by death obtaining among a given number of persons all of whom begin life at the same moment, and it might appear that a probability curve erected on such a basis would misrepresent a curve constructed from the data of an actual population in which at each point of time, during the five year period, infants are being born. But an analysis of the latter data shows that the two situations are similar and that their probability curves are practically identical. For the whole United States population such a curve yields approximately 86 as the average percentage of children, out of all those born during the five years, who are living: conversely the average percentage dying is 14.

But the life table does not distinguish between cities and country and it is most likely that child mortality is somewhat greater in the former than in the latter. Evidently the rural percentage should be less and the urban percentage more than 86. The approximately correct percentage in each case was determined by the following method of calculating rural and urban mortality rates for rural and urban children under 1 year of age, discovering their ratio, and applying it to the five year age group.

Of the Negro population, 2.6 per cent. and of the white, 2.4 per cent., are under 1 year of age.† But the application of 2.4 per cent. to both rural and urban populations does not yield the total number of children of that age. For the national population, also, 9.9 per cent. of the urban and 13.0 per cent. of the rural are under 5 years of age. The same disproportion must be true of children under 1 in the groups, since the ratio of registered deaths in rural districts of the one year and five

*Quarterly Publications, Am. Stat. Assn., Sept., 1914, pp. 228–235.
†Census 1910, Vol. I, p. 314.

Digitized by Google

year age groups and that in urban districts of similar age groups are almost identical, being 72.0 and 71.3, respectively. Hence, with 2.4 per cent. as a basis, multiples which, applied to urban and rural populations, will produce the census number of children under 1 year of age, may be obtained by trying out 2.39, 2.38, etc., for urban and 2.41, 2.42, etc., for rural. It is found that 2.25 applied to the urban population and 2.55 applied to the rural population produce the census number of children under 1, a difference of 6.2 per cent. in favor of the rural. Proceeding in the same nanner relative to Negro infants, the corresponding rates of 2.2 and 2.8 are established.

From the Mortality Report* it is possible to estimate the number of Negro children who survived at the end of a year by taking 2.8 per cent. of the rural and 2.2 per cent. of the urban population, and by the use of the number of deaths reported in each case, estimate the infant death rates. The consequent rural Negro infant death rate is 11.6, and the urban rate is 18.7. The rates established for white infants are: rural, 7.96; urban, 11.9. Weighting the rural rates in each case by their respective white and Negro populations and treating the urban rates in like manner, there ensues a rural infant death rate of 8.5 and an urban rate of 12.3. Should we desire to weight these in turn by their respective populations, a national infant death rate of 10.25 is established.

Granting that the same difference between the rural and urban death rate for children under 5 exists as for those under 1, it is not difficult to determine how the previously obtained 86 per cent., the average percentage of children under 5 who are living out of all who were born, is to be distributed to rural and urban groups. For if $(91.5+87.7)\div 2=89.4$, then $\frac{x+y}{2}=86$. Then $91.5 : 87.7 :: x : y$. Whence x equals approximately 88 and y equals approximately 84. Hence the average percentage of rural children under 5 who survive is 88, that of urban children, 84.

We may proceed now to apply these rates to rural and urban populations. In 1910 there were 4,200,291 urban children under 5 years of age. Since these represent 84 per cent. of

*1911, p. 150.

Digitized by Google

those born, the number of births must have been 5,000,000. Dividing this number by the urban population midway between 1907–1908, a rate of 25.2 is secured as the birth rate. Proceeding to apply the rural percentage, 88, in the same manner, a rural rate of 30.3 is obtained.

As a result of applying 84 per cent. to the urban population of each of the divisions of the United States and 88 per cent. to their rural populations, urban and rural birth rates are established for the nine divisions. These birth rates will be found in columns 1 and 2 of the accompanying table, Table I. Column 3 contains the birth rates of the several division populations. These rates were established by weighting the rural and urban rates with the rural and urban populations of 1900. It is seen that division rural birth rates have a range of over 16, the lowest being 21.0, New England, and the highest, 36.4, West South Central. There is a much smaller range in the urban communities, the Pacific being the lowest, 20.4 and the West South Central the highest, 27.0. As to division rates, the Pacific has the lowest, 23.1, the West South Central the highest, 34.1. To most persons it will be a surprise to learn that the Pacific division has a lower birth rate than New England.

TABLE I.

RATES OF BIRTH, DEATH, AND NATURAL INCREASE IN THE UNITED STATES FOR THE DECADE 1900–1910.

Division.	Birth Rate.				Death Rate.			Rate Natural Increase.		
			Total.							
	Rural.	Urban.	Present Estimate.	Children's Bureau Estimate.	Rural.	Urban.	Total.	Rural.	Urban.	Total.
New England..	21.0	24.4	23.6	24.4	16.0	17.1	16.9	5.0	7.3	6.8
Middle Atlantic	25.5	26.4	26.0	26.2	14.8	16.8	16.2	10.7	9.6	10.4
East North Central.....	25.2	24.7	25.2	24.2	12.8	13.9	13.4	12.4	10.8	11.6
West North Central.....	28.4	22.5	26.3	25.5	10.3	12.4	11.0	18.1	10.1	15.8
South Atlantic.	34.4	25.4	32.1	31.2	15.5	19.4	16.4	18.9	6.0	16.0
East South Central.....	34.4	25.4	32.2	31.6	14.7	18.0	15.3	19.7	7.4	17.8
West South Central.....	36.4	27.0	34.1	32.2	12.5	16.8	13.4	23.9	10.2	21.6
Mountain.....	31.6	25.5	29.4	27.0	10.5	15.0	12.1	21.1	10.5	17.6
Pacific.....	24.6	20.4	23.1	19.9	12.0	13.8	13.0	12.6	6.6	9.8
United States	30.4	25.2	28.6	(27.1)	13.5	16.4	14.7	16.9	8.8	13.65

Digitized by Google

It will be noted that the New England and Middle Atlantic divisions are the only ones in which the rural birth rates are lower than the urban. This is probably due to the large foreign born population, chiefly urban, with its high birth rate. The Southern divisions generally, show a much greater difference in this respect than do any other divisions, the disparity arising from large variations in the rural rates, there being no pronounced fluctuations in the urban rates. The presence of the prolific Negro population in those divisions doubtless accounts for this condition.

With the above division birth rates may be compared the estimates of the United States Children's Bureau.* The fourth column, Table I, contains these estimates. That for the United States in this column results from totaling the number of children reported for the various divisions and dividing by the population of the United States. The two sets of estimates are not far apart, save in the last three divisions. The author's estimate is, however, supported by the following evidence: First, the rate for the nation, 28.0, obtained by weighting the division rural and urban rates is almost identical with the national rate, 28.1, secured by obtaining separate birth rates on the basis of the total number of rural and urban children under 5 in the whole nation and proceeding to weight them. Second, these rates agree with that obtained by treating the children of the nation, irrespective of divisions, and of the rural and urban communities. Third, the rate of natural increase for the United States as determined in the earlier part of this paper, is almost identical with that obtained by the use of the birth and death rates presented in Table I.

As has been indicated, the mortality rates contained in the United States Mortality Reports, cover only about two thirds of the national population, none being available for many states and for one entire division. Yet the distribution of the area from which reports are made is such that a mortality rate may be approximated for neighboring and similar areas. Space does not permit of giving the reasons for my estimates for the various states that do not report on mortality. Suffice to say, I have given each state careful attention.

*Birth Registration, Monograph No. 1, 3d edition, 1914, p. 18.

Digitized by Google

The mortality estimates in Table I were made on the basis of the data given in the Statistical Abstract of the United States.* Column 5 presents the simple average of published and approximated mortality rates of the states as urban rates by divisions, and column 6 contains the same results for urban communities by divisions. The rates presented in column 7 were obtained by weighting the rural and urban rates in columns 5 and 6. The national rural and urban death rates are the average for the five years, 1906–1910, that for the nation as a whole represents the weighting of all the division rural and urban rates.

The rates for the West North Central and the Mountain divisions seem low. It is possible that they are too low, but it is to be remembered that it is altogether probable that both the climatic conditions and the age distribution of the populations of these divisions are favorable to the existence of low rates. The reports of mortality from such states as make reports may all be wrong, but the reports we get certainly indicate that the conditions of life in those sections favor longevity.

Columns 8, 9, and 10 of Table I, embody the results, for most part, derived from a treatment of the rates contained in the columns under Birth Rate and Death Rate. The rural column exhibits a relatively steady increase of rates, beginning with New England and proceeding toward the Pacific. A vast fluctuation in these rates is obvious, those of the West South Central and Mountain divisions being from four to five times the New England rural rate. It is seen, that should the indicated rate persist, the West South Central division would almost double its population in four decades by natural increase alone. In the same manner it would require New England 200 years to accomplish that result.

The urban column offers its surprises in that three other divisions show as low a rate of natural increase as either New England or the Middle Atlantic. Since the Negro population in the South is largely a rural population, it is plain that the native white stock of the southeastern divisions are multiplying about as slowly as is that of rural New England. The native white stock of urban New England is doubtless in the

*1913, pp. 74–75.

Digitized by Google

same or even in a less fecund condition, since its urban rate is largely sustained by the more prolific immigrant stock. It is also noteworthy that the Pacific urban rate is less than that of New England. This is partly due to the fact that the population of this division contains an unusually large percentage of individuals from 25 to 44 years of age and of unmarried males and females. It is probable also that the native stock has approximately the same tendency toward infecundity exhibited in New England and the southeastern cities.

The national rural and urban rates in each instance are derived in the same manner as are the division rural and urban rates, by taking the difference between the national birth and death rates. It certainly is a matter of surprise that the rural rate is almost double the urban rate. The assumption which is found repeatedly in published works is that the urban birth rate exceeds the rural rate, and the conclusion is usually drawn that the natural increase of the former is consequently greater.

The total rates of natural increase are the weighted averages of the rural and urban rates, division by division. As in the case of the rural column, a progression is noticed in the size of the rates from New England down to the West South Central division, a decline occurring after that. The Pacific rate approximates that of the Middle Atlantic. That of the West South Central division is more than three times that of New England and over twice that of the Middle Atlantic or the Pacific. Those who believe in economic determinism will find comfort in this column, for it is apparent that the rate of natural increase largely follows the pressure or freedom from pressure of economic opportunities. However, a second thought reveals the fact that race and standards of culture play a large rôle, because the highest rates are in the division that include the bulk of the Negroes, where heavy breeding occurs in spite of their backward economic conditions.

When the results embodied in the last three columns of Table I, especially those of the rural and urban rates, were obtained, the writer was inclined to think a huge mistake had been made. But a careful and repeated reëxamination of the data and processes on which the results depend compel him to think the findings are essentially true. Should it be said that

Digitized by Google

the birth rates are too high or too low, the reply is that they must be so for all of the divisions, for the same method was applied to all data alike. And should it be said that the mortality rates for certain of the divisions are too low, the answer is that then they are too low alike for the rural and urban communities of the divisions, for the same method was applied in every case. The result would be that there might be a small modification of urban and rural death rates in some divisions, and a slight lessening of the natural increase rates in general due to a uniform reduction of the birth rates. But the results would not be changed greatly. Consequently, Table I gives a comprehensive and, I believe, a fairly accurate picture of the situation for the United States as a whole, for the several divisions, and for rural and urban communities as regards the vital facts of natural increase.

In concluding this treatment of natural increase, I desire to remark on the close agreement existing between the rate of natural increase, 13.65, to be found at the foot of column 10, Table I, and that of 13.7 obtained by an analysis of the population increase between 1900 and 1910 into its factors of emigration, immigration, and native born. The latter percentage, 13.7, was first obtained and the fact that the former percentage was so nearly identical was regarded as an indication that the methods by which it was secured were reliable. To the writer, this agreement is no small evidence that the birth rates and the death rates contained in Table I are approximately correct. Were they considerably modified the resulting rate of natural increase for the nation would, by the amount of the modification, fail to agree with the first established rate.

One of the most interesting undertakings in the field of the statistics of population is that which deals with the factors that account for growth of population and seeks to accord to each one its relative force in producing the change. Little of a scientific analytical character has been done toward the accomplishment of this object and the territory is practically a virgin one. We know relatively nothing in an exact way regarding the forces at work in the population of our nation until this task is accomplished, and the hope of contributing something that will serve as a foundation for further attain-

Digitized by Google

ments in that direction is a sufficient spur to drive a worker into this complicated and difficult field.

The elements that enter into the rural and urban changes of our national population in the direction of determining its growth are: natural increase, emigration, immigration, incorporation, and rural migration.

In round numbers the increase of urban population during the last census decade was 11,826,000. This is to be credited to incorporation, immigration, and natural increase. Had it not been for emigration the gain would have been larger. This emigration from urban districts amounted to 294,000, therefore the potential urban increase for the decade totals 12,120,000.

According to the Census, incorporation means three things: the addition of suburban territory to urban districts; the growth of rural villages past the 2,500 population point, the dividing line between rural and urban communities; and the falling back of some urban communities into the rural class. During the decade, ending 1910, approximately 813,000 inhabitants passed over into the urban group according to the census figures. Computing the increase of this population by applying the national natural increase rate already established, the final urban increment by incorporation is found to be 924,000, or 7.8 per cent. of the urban increase.

Immigration is the largest factor in accounting for the increase of urban population. The foreign born population of the United States in 1900 was 10,341,276; in 1910, 13,515,886. In 1900, 66.8 per cent. of the white foreign born population was urban, and in 1910, 72.2 per cent. Consequently the urban foreign born in 1900 must have been 6,908,000 and that of 1910, 9,745,000. To the former number belongs also that portion of the foreign born rural population which became urban by 1910, a sum of 111,000, swelling the 1900 urban foreign born population to 7,019,000. Since, as we have seen, but 80 per cent. of this population remains by 1910, or 5,615,000 the net immigration for the decade was 4,130,000. Increasing this sum by 17.4 per cent., the rate formerly established to represent the decennial natural increase of immigrants, it is found that the national urban population was enlarged by

Digitized by Google

reason of immigration, by 4,849,000, or 41 per cent. of the total urban gain.

How far is the urban gain in population to be accounted for by the natural increase of the population resident in urban districts in 1900? Our urban rate of natural increase has been established during the preceding treatment (See Table I, column 9). Applying this rate, 8.8 per cent., to the urban population of 1900, 30,797,000, gives 2,710,000 as the amount of such natural increase. But since urban populations sustained a decennial loss of 294,000 by reason of emigration and since, it may be assumed, this loss was sustained chiefly by the population that was present in 1900, the amount of natural increase may be diminished by the amount of emigration, leaving a balance of 2,426,000 to be assigned to natural increase. This represents 20.5 per cent. of the urban decennial gain.

It is now possible to estimate the amount of rural immigration by taking the difference between the sum of the ascertained factors and the decennial urban increase. In round numbers this amount is 3,637,000, or 30.7 per cent. Hence we may gather the results in Table II:

TABLE II.

FACTORS OF URBAN INCREASE FOR THE UNITED STATES AS A WHOLE, 1900-1910.

Factor.	Amount.	Per Cent. of Urban Increase.
Incorporation	924,000	7.8
Immigration	4,849,000	41.0
Natural increase	2,426,000	20.5
Rural migration	3,637,000	30.7
Total	11,826,000	100.0

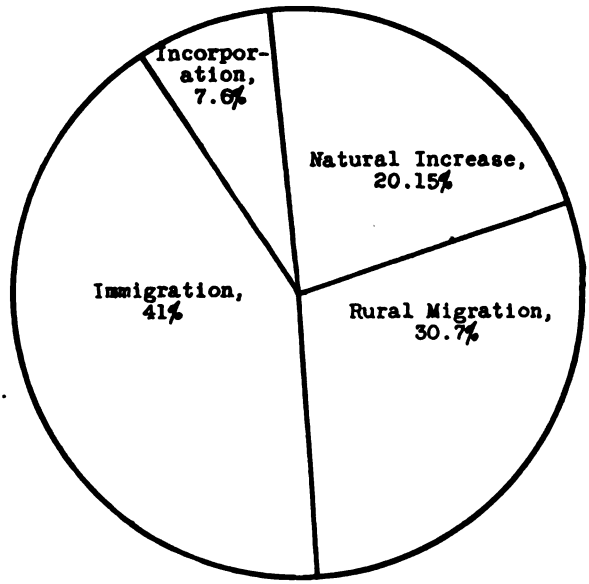
The results of this table may be presented in graphic form.

The proportion of these urban factors to each other differs but slightly from that given in the published estimate* previously referred to. But my figures vary widely from those given by Mr. Earl Clark in his paper "Contributions to Urban Growth."† Mr. Clark did not seek, however, to estimate the amount net immigration is swelled by its own natural increase which is found to represent over 718,000 persons. Neither

*See QUARTERLY PUBLICATIONS of the Am. Stat. Assn., Sept., 1915, p. 649.
†*Ibid.*, p. 671.

Digitized by Google

GRAPH II. RELATIVE FORCE OF FACTORS DETERMINING URBAN INCREASE IN THE UNITED STATES, 1900-1910.



did he deal with the factor of incorporation which we have seen affects the situation in various ways, although he excludes the incorporated population from his total urban increase. It may be said further, that he did not have the advantage of the detailed results of the work on natural increase of population which are presented in this paper.

It will not be necessary to indicate the detailed calculations made in arriving at the results representing the factors concerned in the rural increase for the United States during the decade. The methods and processes were in general similar to those involved in the urban analysis. It is well to note, however, that both incorporation and rural migration in Table II represents rural losses, in addition to the factor, emigration. Thus, to secure rural migration, from the sum of immigration and natural increase is taken the sum of incorporation, emigration, and census estimate of rural increase.

These amounts are as follows: Immigration, 1,290,000; natural increase at rate 16.9, 7,638,000; total gain in popu-

Digitized by Google

lation, 8,928,000. From this is taken: Census increase, 4,151,000*; incorporation, 924,000; emigration, 294,000; total, 5,369,000. The difference, which represents rural migration, is 3,559,000.

The results of rural analysis for the nation as a whole are embodied in the following table:

TABLE III.

FACTORS OF RURAL INCREASE FOR THE UNITED STATES AS A WHOLE, 1900-1910.

Factors.	Amount.	Per Cent. of Census Increase.
Immigration	1,290,000	31.0
Natural increase	2,861,000	69.0
Total	4,151,000	100.0

It is observed that only about 37.5 per cent. of natural increase of rural communities is devoted to accounting for the Census increase, the remainder being accounted for under the three factors of loss: emigration, incorporation, and rural migration. It must also be noted that the total rural increase is not the Census statement of increase but an amount derived by subtracting the rural population of 1900 living in territory that was rural in 1900 from the rural population living in rural territory in 1910. This is made necessary to permit the use of the factor of incorporation in estimating the force of rural migration. A difference of over 80,000 appears between the estimates of rural migration made for rural and urban populations. We shall find further variations from both of these estimates, as for the other factors, when we arrive at a closer analysis of rural and urban populations by divisions. Such variations are to be expected because of some uncertain elements in the computations, but it is hoped that they do not greatly militate against the value of the estimates.

It might seem a perfectly simple matter, since the force of each of the various factors has been determined for the nation relative to urban and rural districts, to apply the results to the divisions. But a consideration of a single division reveals new difficulties. Thus, how is emigration to be distributed

*Abstract 13th Census, 1910, p. 56.

Digitized by Google

among the several divisions so that each shall have its actual amount? And what is to be done with the new element in the situation, namely, the interdivision migration of native born persons?

To treat in detail each division, and to do the same for the urban and rural districts in each division, would require too much space. We shall have to be satisfied with a somewhat syncopated handling of a single division and the presentation of the results from the other divisions in a comprehensive table. Divisions as a whole will first be considered; then a treatment of their urban and rural districts will be undertaken.

I have selected the Middle Atlantic division to illustrate the method of making estimates. Any other division would do equally well, since a uniform method is employed. The force of immigration is ascertained exactly as it was for the nation as a whole, and for the national rural and urban districts. The net immigration for the decade so ascertained is 2,197,466, its natural increase at 17.4 per cent. is 382,000, giving a total immigrant contribution beyond the population of 1900 of 2,579,000. Since the increase of population for the division as recorded by the Census was 3,861,214, immigration with its natural increase constitutes about 67.0 per cent. of such increase.

Since emigration enters into the calculation of the potency for natural increase of the 1900 population it must be regarded. Had we to consider only accounting for the Census amount of increase it might be disregarded. Its consideration also is demanded as a foundation for calculations relative to the factor of rural migration to be determined hereafter.

There is no certain indication for distributing the effect of emigration among the divisions. One might expect that the drain from some divisions would be proportionately larger than from others. I have made calculations for all entire divisions, as well as for their respective separate rural and urban parts on the basis of several different assumptions. The criterion of availability applied to these methods, as also of those regarding the various modes of treating interdivision migration of native born Americans, has been their effect on the amount of rural migration as separately determined for

Digitized by Google

rural and urban regions (Tables II and III). In each case, I have concluded that the one which secured most nearly identical sums of rural migration in the respective divisions when determined by the two processes of analysis, rural and urban, was the one most nearly correct. As a consequence I have returned to my initial plan of distributing the effects of emigration among the divisions in proportion to their population, and within divisions between urban and rural according to their respective populations.

In 1910 the rural and urban populations of the United States were so nearly equal that the sum of emigration was halved and an equal portion allotted to city and country alike. Since the total amount, not increased by natural gains, was estimated at 550,000, urban and rural regions would each suffer a loss of 275,000. Middle Atlantic urban population in 1910 formed 32.2 per cent. of the total national urban; hence its allowance of emigration was 88,500. In like manner rural Middle Atlantic, since its population was 11.3 per cent. of the national rural, was assigned 31,050. Since the decennial urban rate of natural increase is 9.6 per cent. (Table I) and the rural 10.7, the assigned sums are increased to 93,400 and 32,650, respectively. The urban amount, 93,400, constitutes approximately 2.4 per cent. of the Census increase of the division.

According to the Census report* Middle Atlantic, in 1910, had 932,467 more persons born within the division who lived outside it than it had native born of other divisions living within itself. But compared with the record of 1900, when the difference against Middle Atlantic was 1,088,150, the division really gained in native born; that is, the division was 155,683 better off in 1910 than in 1900. The first assumption likely would be that the latter amount is the sum we need and represents a gain in population. But, as is the case with other factors, logical analysis of the situation is found to be necessary in order that the true force of the migration of this element may be discovered.

Each division relative to the native born migrant must be viewed and treated just as the nation is viewed and treated

*13th Census, Vol. I, p. 693.

Digitized by Google

relative to foreign born. There is an inflow and an outflow, births and deaths, and these modifying factors must be measured. Hence a double analysis must be made; one for those born in the division but living outside for 1900 and 1910, and one for those born outside but living within for both dates.*

For Middle Atlantic in 1900, persons born in the division but living outside numbered 1,808,060. These would suffer inroads by death during the decade, and I have used the national decennial death rate of 15 per cent. for all divisions. I consider this better than the separate division mortality rates, since there is no way of knowing what length of time the migrants have been acted on by the local rate. Consequently only 1,538,000 are living by 1910. In 1910, those who are born in the division but live outside it number 1,881,406. The difference, 343,400, between this sum and 1,538,000 represents the gain in the number living outside the division.

In the same manner, persons born outside Middle Atlantic living inside in 1900 numbered 719,910 of whom 620,000 are living in 1910. Persons of the same character at the 1910 Census number 948,939, the difference between this and the survivors representing a net gain of 329,000.

These amounts, 343,400 and 329,000, are also influenced by the decennial rate of natural increase. For the same reasons given for the use of the national instead of the local death rate, I chose the national rate of natural increase, 13.7. Applying this rate to each amount, they become 367,800 and 352,000 respectively, wherefore their difference, 15,000, represents the loss of native born to other divisions Middle Atlantic sustains. Hence what seemed to have been a great loss, the Census 155,683, by analysis of the situation appears as a small one. And the same is true in the case of all eastern divisions whose census record appears positive.

Having ascertained the value of each of the factors, immigration, emigration, and migration of native born, it is possible to discover the amount of natural increase by a new method. Since the latter two factors represent losses of the 1900 population, they must be added to the Census increase of 1910 to discover what the full potency of the former population act-

*For Census data, 13th Census, Vol. I, p. 692.

Digitized by Google

ually was, giving a total sum of 4,002,264. From this is taken immigration, 2,579,000, leaving the amount of natural increase, 1,423,264. Dividing the latter amount by the population of 1900, a rate of natural increase of 9.2 for the division is obtained. By our other method of securing the division rate of natural increase, a rate for the Middle Atlantic division of 10.4 was established.

When the same method is applied to each of the other divisions as was just applied to the Middle Atlantic, the data to be found assembled in Table IV are obtained.

Relative to this table of results some remarks are pertinent. Lest there be a misunderstanding as to what the totals imply, it is sufficient to say, first, that column 5 plus column 7 less column 2 should equal the total in column 1. The difference is about 23,000. Second, the total in column 2 should equal the original total estimated for the United States, 588,000. The difference is seen to be slight. Again columns 3 and 4 should balance, but the difference is only about 5,000. Finally, the force of immigration as originally estimated for the nation was 6,159,000, as compared with the total of 6,153,000 of column 5. A portion of these small differences no doubt are due to the use of the slide rule in most of the hundreds of computations involved in evolving the table. Some of the differences are evidently due to the application of uniform rates of death and natural increase to widely differing populations, as in the case of immigration and of the native born migrants. But considering the complexity involved, it is rather remarkable that the divergences are not greater. For instance, to secure the amount of natural increase for the several divisions and the nation involved the treatment of the new factor, native born migration, yet the difference between the total so secured, 10,429,000, differs from the original estimate, 10,412,000, by only 17,000.

Only three divisions, all western, add to their population by means of an actual excess of income over outgo of native born Americans. Column 6 shows that immigration contributes slightly to the increase in the South and Southeast but that its force in that direction is preponderent in the North and Northeast.

Digitized by Google

TABLE IV.

FACTORS ACCOUNTING FOR INCREASE OF POPULATION BY DIVISIONS, 1900–1910.

Division.	Increase According to Census.	Loss.		Gain.								
		Emigration: Number with Increase.	Migration Native Born: Number with Increase.		Immigration.		Natural Increase.					
					Number with Increase.	Per Cent. of Census Increase.	Amount.	Rate.	Per Cent. of Census Increase.	Actual Contribution to Census Increase.	
										Amount.	Rate.
New England	960,664	42,830	22,000	.....	784,000	81.5	241,494	4.3	25.0	176,664	17.3
Middle Atlantic	3,861,214	126,050	15,000	.....	2,579,000	66.7	1,423,364	9.2	36.9	1,282,214	33.2
East North Central	2,265,040	116,400	688,000	.....	1,142,500	50.0	1,927,200	12.0	85.0	1,122,460	50.0
West North Central	1,290,500	73,400	658,000	.....	457,906	35.5	1,563,900	15.2	121.0	832,594	64.5
South Atlantic	1,751,415	75,840	104,000	.....	149,270	8.0	1,781,985	17.1	102.0	1,602,145	91.0
East South Central	862,144	53,980	457,700	.....	18,050	2.0	1,355,800	17.0	157.0	840,094	97.4
West South Central	2,252,244	55,500	.....	470,000	162,622	7.2	1,675,000	25.5	74.0	1,619,600	72.0
Mountain	958,860	16,700	.....	450,800	248,597	26.0	276,160	16.5	29.0	259,400	27.0
Pacific	1,755,612	26,300	.....	1,006,000	609,827	34.5	186,700	7.7	10.6	160,000	9.0
Total	15,977,669	581,600	1,937,000	1,932,000	6,153,000	....	10,429,000	....	.....	.....	....

Digitized by Google

Column 8 gives the rates of natural increase established by the method of division analysis. In general, they confirm the findings in Table I, column 10. In four divisions the difference of rates is less than 1 per cent., in two others slightly more than 1, in two others between 2 and 2.5, and in one 3.9. Nor is one set of rates uniformly higher than the other. Thus the rates in Table I exceed those in Table IV only in the New England, Middle Atlantic, West North Central, Mountain, and Pacific divisions. The weighted average for all the divisions gives practically the same rate of natural increase for the United States as obtained by the other method, 13.7 vs. 13.65, and agrees with the rate previously established for the nation.

The difference between the two sets of rates may be due, in New England, to the fact that the estimated rate of natural increase for immigration is larger than conditions of life in that division justify. If this assumption is true, the new rate of natural increase for the division population would be increased by that much. The very large rate in the West South Central division may be due to applying the national mortality rate to the native born migrants treated relative to that division, when the local death rate is higher and should have some influence in the estimate. However, I know of no means of taking such considerations into account.

The actual contribution to the division increase as recorded in the Census, given in column 10, differs from the amount of natural increase. In the first six divisions this contributive amount is ascertained by obtaining the difference between the Census increase and immigration; in the last three divisions, by finding the difference between the Census increase and the sum of immigration and native born migration. In the case of the East South Central division it is apparent that the Census increase would have been over one and a half times its actual amount had the native born loss not occurred.

The factors which enter into the determination of the urban increase in each of the various divisions are those discovered in making calculations for each division as a whole, together with the new elements of incorporation and rural migration, both of which entered into the analysis of the national increase with respect to urban and rural. The force of the various

Digitized by Google

TABLE V.

URBAN INCREASE IN THE UNITED STATES, 1900–1910.

Division.	Increase According to Census.	Factors Affecting Urban Increase.															
		Losses.		Gains.													
		Emigration.	Interdivision Migration of Native Born.			Immigration.		Incorporation.		Natural Increase.					Rural Migration.		
													Contribution to Census Increase.				
			Amount (Loss).	Amount (Gain).	Per Cent. Census Increase.	Amount.	Per Cent. Census Increase.	Amount.	Per Cent. Census Increase.	Amount.	At Rate.	Per Cent. Census Increase.	Amount.	Per Cent.	Amount.	Per Cent. Census Increase.
New England	985,166	32,150	18,160	.....	....	742,000	75.4	20,800	2.2	210,000	4.7	21.3	160,000	16.2	62,200	6.3
Middle Atlantic	3,288,600	92,800	10,620	.....	....	2,207,000	61.0	254,000	7.2	326,000	7.3	33.0	213,000	21.5	(a)53,800	0
East North Central	2,397,000	65,300	362,000	.....	....	1,005,200	42.0	182,000	7.2	975,000	9.6	23.8	513,000	15.6	314,579	8.7
West North Central	927,172	26,400	221,500	.....	....	261,640	28.0	83,820	9.0	780,000	10.8	13.3	342,700	14.2	857,100	35.8
South Atlantic	859,521	20,640	27,100	.....	....	90,000	10.5	111,300	13.5	297,600	10.1	32.0	49,460	5.3	532,252	57.0
East South Central	443,173	12,180	83,600	.....	....	14,100	0.6	59,200	13.4	134,000	6.0	15.6	186,060	21.5	572,161	66.5
West South Central	900,259	13,450	.....	104,800	11.6	70,700	7.9	110,200	13.0	83,600	7.4	18.8	(a)14,452	(a)3.3	382,053	86.3
Mountain	406,148	6,370	.....	162,200	40.0	102,500	25.3	37,600	9.2	108,000	10.2	12.0	118,550	13.2	519,809	54.0
Pacific	1,260,000	15,900	.....	572,000	45.5	392,500	31.0	62,300	5.0	57,000	10.5	13.8	50,630	12.5	53,218	13.1
										74,200	6.6	5.8	58,300	4.6	174,900	13.9
Totals	11,826,198	285,190	839,000	722,000	....	4,885,640	....	921,220	....	2,721,400 or 2,837,400	....	....	1,464,248 or 1,517,248	....	3,468,272 or 3,405,472	....

(a) Loss.

Digitized by Google

factors was reckoned much as was the rural and urban gains for the nation as a whole.

The results established by the analysis of these factors determining urban increase in each of the nine divisions have been brought together in the preceding table.

The amount of emigration and of native born migrants were obtained by apportioning the division total in each case (Table V, columns 2, 3, and 4) between the rural and urban districts of the division according to their population. As in the distribution of emigration among the divisions there seems to be no other principle to use in the apportionment. The force of immigration is obtained by the treatment of the urban foreign-born of 1900 and 1910 according to the method previously developed. The factor of incorporation calls for no explanation, except that it represents the population in the division that became urban between 1900 and 1910,* and that since it is involved in the process of reckoning natural increase, according to the meaning adopted in this paper, it cannot be disregarded. The amount of natural increase is found by the use of the division rate for urban population (Table I, column 8). It is impossible to determine it by a process of exclusion as was done in estimating the amount of rural migration. Thus, in the Middle Atlantic division, rural migration equals the difference between the sums involved in columns 1 to 3 and 6 to 10, Table V.

New England gives some difficulty respecting its rural migration. If the process applied to the Middle Atlantic division is employed, it is found that not only is there no population left to rural migration, but at the urban rate of increase for the division, 7.3 per cent. per decade, there is a deficit of nearly 54,000. A similar deficit occurs when the same operation is performed for the New England division as a whole. Evidently, local conditions make some of the rates established inapplicable unless qualified. Which of the factors is overrated, or whether several are involved, there is no cue to determine.

An attempt has been made to estimate what proportion of the urban increase of the several divisions is to be assigned to

*Abstract 13th Census, pp. 56–57.

Digitized by Google

each factor helping toward the gain. The results uniformly occur in the columns headed "per cent. of census increase." It will be noted that columns 13 and 14 estimate the contributive power of natural increase. The same might have been undertaken for any of the other factors of increase, instead, but it is more appropriate for natural increase, since it aids us to see how far natural increase is being displaced by other sources of population. In the case of New England, two estimates are given for natural increase and rural migration; one based on the established urban rate for the division; the other, on a rate obtained by proportioning the division rate of natural increase (Table I, column 10) to the urban part of the division population. With the smaller rate, not only are the amount and contributive force of natural increase enlarged, but the deficit in rural migration (column 15) is turned into a surplus.

Should a comparison be made between the totals for columns 2, 6, 8, and 10, with the estimates for each of the items made previously in this paper, it would be found that some variations from those estimates appear. Using round numbers for the previous and the present computations in their order the estimates are: immigration, 4,885,000 vs. 4,849,000; emigration, 285,000 vs. 294,000; incorporation, 922,000 vs. 924,000; natural increase, 2,721,000 or 2,837,000 vs. 2,710,000; and rural migration, 3,405,000 or 3,468,000 vs. 3,637,000 or 3,559,000. These variations are partly due, presumably, to the use of the slide rule and to local conditions which refuse to conform to general formula, chiefly the latter. The gains and losses due to interdivision migration of native born should balance, but it is observed that they are about 84,000 apart. It is obviously possible that apportioning this factor to urban and rural communities according to population fails to meet the exact needs of the situation.

In using the foreign born population of 1900 and 1910 in order to estimate the net immigration for the decade, it was necessary to reckon the total number from the number of foreign born whites given in the 13th Census.*

Table VI, which presents the results of a series of estimates to discover the force of the various factors determining the

*See Census 1910, Vol. I, pp. 175, 189, and 781.

Digitized by Google

TABLE VI.

RURAL INCREASE IN THE UNITED STATES, BY DIVISIONS, 1900–1910.

Division.	Increase According to Census.	Factors Affecting Rural Increase.												
		Gains.									Losses.			
		Immigration.		Natural Increase.					Migration Native Born.					
		Amount.	Per Cent. Census Increase.	Amount.	At Rate.	Per Cent. Census Increase.	Contribution to Census Increase.		Amount (Gain).	Per Cent. Census Increase.	Amount. (Loss).	Emigration.	Incorporation.	Rural Migration.
							Amount.	Per Cent.						
New England	(a)24,502	42,000	.....	35,920	3.2	.....	.....	....	.....	.....	1,840	6,330	20,800	44,628
Middle Atlantic	213,724	371,659	17.4	56,100	5.0	.....	.....	....	.....	.....	4,380	32,650	254,000	24,448
East North Central	(a)132,256	137,338	.....	550,729	10.7	260.0	.....	....	.....	.....	4,380	32,650	254,000	417,634
West North Central	363,326	196,199	54.0	1,086,984	12.4	.....	.....	....	.....	.....	326,000	51,200	182,000	797,378
South Atlantic	891,894	59,270	6.6	1,340,000	18.1	368.0	.....	....	.....	.....	439,000	47,000	83,820	602,454
East South Central	418,971	3,950	0.9	1,532,034	18.9	172.0	.....	....	.....	.....	76,900	55,200	111,300	456,610
West South Central	1,351,985	91,922	6.8	1,253,117	19.7	298.0	.....	....	.....	.....	373,400	41,800	59,200	363,696
Mountain	452,712	146,097	32.0	1,283,669	23.9	95.0	.....	....	365,200	27.0	.....	42,500	117,500	228,806
Pacific	515,639	217,327	42.0	231,889	21.1	51.0	.....	....	288,600	64.0	.....	10,330	37,600	165,944
				155,736	12.6	30.0	.....	....	434,000	84.0	.....	10,400	62,300	218,714
Totals	4,208,251	1,265,760	.....	7,470,078	....	.....	.....	....	1,087,800	.....	1,221,520	297,410	928,520	3,295,264

(a) Loss.

Digitized by Google

growth of rural population, calls for little comment. As in the case of the urban total, I have made two estimates of the amount of natural increase, and of rural migration, for New England, based on the two different rates. The employment of the lesser rate in both urban and rural analysis gives fairly similar amounts of rural migration.

The sums expressing rural migration in Tables V and VI for the several divisions should, of course, be identical. In some cases they are very wide apart. I have worked out rural migration on the basis of several different kinds of modifications of certain factors. Thus I distributed emigration differently, assigning a greater share to certain divisions, and to the urban or to the rural in given divisions, than in the tables. I also tried interdivisional migration of native born according to the untreated census figures and according to the different rates of natural increase and of death. But the present plan has more nearly reconciled the two series of amounts of rural migration.

Column 6 of Table VI is especially interesting. It is observed that in three divisions the natural increase amounts to several times the actual division increase, that in others it is equal to or greater, and that in two the Census expresses a deficit, in one of which cases the amount of natural increase is over one million. This situation, of course, signifies that the other sources of population have been supplanting individuals born in the division or that the latter have gone to other divisions.

A comparison of the totals of Table VI with similar amounts previously estimated, as in the case of the figures in Table III, shows these differences: For immigration, 15,000; natural increase, 168,000; emigration, 3,000; incorporation, 4,500; rural migration, 304,000. The total for rural migration differs by 100,000 to 150,000 from those given in Table V, column 15. The division losses and gains also show a variation in totals of about 135,000. In a field of such extreme complexity, however, and where there is no known means of eliminating assumptions, the occurrence of divergences would be anticipated by the statistician.

Digitized by Google

A comparison of the estimates of rural migration and of interdivision migration of native born secured by the two sets of analyses which were made, the one rural and the other urban, is of assistance in explaining how the apportionment of the results of migration of native born between urban and rural communities may produce great fluctuations. For this purpose it is useful to draw those data together in a table.

TABLE VII.

RURAL AND URBAN AMOUNTS OF RURAL MIGRATION AND INTERDIVISION MIGRATION OF NATIVE BORN, BY DIVISIONS.

Division.	Rural Migration.		Interdivision Migration Native Born.			
			By Rural Analysis.		By Urban Analysis.	
	By Rural Analysis.	By Urban Analysis.	Loss.	Gain.	Loss.	Gain.
	44,628	62,200	.....	.....	.....	.....
New England	24,448	(a) 53,800	1,840	.....	18,160	.....
Middle Atlantic	417,634	314,579	4,380	.....	10,620	.....
East North Central	797,378	857,100	326,000	.....	362,000	.....
West North Central	602,454	532,252	439,000	.....	221,500	.....
South Atlantic	456,010	572,161	76,900	.....	27,100	.....
East South Central	363,696	382,053	373,400	.....	83,600	.....
West South Central	228,808	519,809	.....	265,200	.....	104,800
Mountain	165,944	53,218	.....	288,600	.....	162,200
Pacific	218,714	174,900	.....	434,000	.....	572,000
Total	3,295,264 or 3,275,086	3,468,272 or 3,405,472	1,087,000	1,121,520	839,000	722,000

(a) Deficit.

An inspection of the data in Table VII is sufficient to indicate that the differences between the amounts of rural migration in the first two columns might readily be accounted for on the supposition that the effects of interdivision migration of native born are not truly apportioned between rural and urban communities. Thus, in the West South Central division, the gains, rural and urban, if differently distributed between those districts, could easily reconcile the differences in the amount of rural migration given by rural and urban analysis. In the first two divisions, this is not so true. But a different apportionment of emigration losses would be sufficient to reconcile the divergencies there. These estimates must remain far apart, however, until some method has been

Digitized by Google

found to apportion the results of emigration and migration of native born according to the actual events. Meanwhile it will be best to state the amount of rural migration occurring in the different divisions in terms of the minimum and maximum furnished in columns one and two of Table VII.

Relative to the rural migration of the nation as a whole, we have found a range of estimates from 3,275,000 (Table VII, column 1), which was the lowest estimate given by rural division analysis, to 3,637,000 (Table II), the maximum, given by the analysis of the whole urban population of the nation. This is a range of 362,000, or about 10 per cent. of the maximum estimate. Perhaps the true estimate, if known, would fall somewhere between the two, approximating 3,500,000.

Digitized by Google

FORTY-FIVE YEARS OF U. S. NAVAL MORBIDITY STATISTICS OF DEGENERATIVE DISORDERS.

By Charles N. Fiske, M.D., *Surgeon, U. S. Navy.*

It might well be an old wag to say that the navy has had a floating population; ethnically this fact is as important as that its personnel is derived from the civil population; any observations concerning its inherent traits reflect with considerable accuracy corresponding conditions among civilians of similar sex, age, physique, and intellect. The word "accuracy" is used advisedly for the reason that these observations are unusually trustworthy from being made by specially trained officers who maintain constant relations with their clientele; this is particularly true of naval medical officers whose association with the navy personnel is continuous as compared with the casual observations of civilians who rarely consult their physicians so long as they consider their health normal. Moreover the military services are practically the only groups of our population concerning whom complete morbidity statistics have been published; their officers and men are undoubtedly the only large community of whom fairly frequent periodical physical examinations have been required and in whom the movement of disease has been recorded continuously for 45 years.

Standards of age, physique, and intellect for enlisted men who have constituted over 90 per cent. of the navy and marine corps have changed perceptibly but not materially for present considerations; for while the average age has been reduced and physical and intellectual requirements have certainly risen, the proportion of "continuous service" men as shown by percentage of reënlistments has probably increased sufficiently to offset any undue advantage attributable to more discriminating selection at recruiting stations in recent years.

Statements, more or less arbitrary, are becoming commonplace, to the effect that the degenerative diseases of mature age, such as the scleroses and malignant proliferations, are insidiously although steadily increasing among the general

Digitized by Google

population while preventive medicine is occupied with successfully diminishing morbidity and mortality from the exanthemata and infections of youth, the inference being that sanitation and hygiene should begin to be more concerned with the former. It may well be doubted if these familiar ideas, which already approximate the category of platitudes, are based upon precise data or even tenable premises. Let it be recalled that 25 years ago when appendicitis became the vogue it was generally questioned if the disease was at all prevalent during the preceding generations and it was only after extended delving into old case histories and supposedly antiquated medical literature that it was found to be by no means a new entity. Standards of living change; published intelligence increases; precision in diagnosis certainly improves, although yet remote from perfection; disease terminology and classification advance; hospital clinics serve classes of people varying from one decade to the succeeding; and surgical practice, medical art, and even the so called pure sciences can not be acquitted of insecure impulse and bias. Many such and other tendencies occur to invite pause when one is told that cancer is alarmingly on the increase or that the strenuous high tension, under which many (but not necessarily average) Americans are pursuing dollars, is bringing the nation to doom through Bright's disease and arterio sclerosis, when the influence of the compound microscope, Virchow's cellular pathology, and the sphygmomanometer may have been the insidious factors.

In the first place does it always occur to the fluent speaker and facile writer of medical altruism that, as Sir Victor Horsley believed, we have been sparing the weak and unfit in youth to furnish material for degenerative disorders in maturity thereby attempting to defeat Nature's provision for the survival of the fittest? Over how many decades may a given population continue to boast of a death rate of 12 to 15 per thousand if the mathematical absurdities involved are to be judged in the light of human experience?

The current generation of health publicists, while admitting that a life saved in youth must of necessity cease from some cause in later years, may preach that disorders of age are or

Digitized by Google

should be as preventable as those of youth, but it will take at least the next generation to preach the hope that humans need never die at all.

The tendencies which prompt such hypotheses should be susceptible of demonstration through very interesting studies of mortality rates in definite age groups for special diseases of the registration area published by the Vital Statistics Division of the United States Census, which constitute "another story" and are not practicable in this paper.

The 10,000 to 70,000 males, aged 16 to 64 years, in the United States Navy and Marine Corps throughout 45 years are conceded to be all too few upon which to base any dogmatic conclusions in this direction, yet the verdict of "not yet proved" which haunts the minds of some has prompted a study of the navy's morbidity statistics to demonstrate a tendency to confirm or deny the impressions already stated as loosely thought facts. The tabulation will occupy several weeks of spare hours and one hopes that, whatever tendency may be declared in the result, it will possess some value and not be in vain.

If it be granted that an unbiased analysis of the movement of so called degenerative diseases among the naval personnel appointed and recruited from all geographical sections of the nation's civil population would be worth while, the first step must be to determine what classes and entities of disease merit inclusion and what factoral value shall be assigned to cases, invalidings, and deaths.

During the present troublous times a writer has to exhibit proof of his neutrality, and then anyone but himself may still question it. One can confess to such preconceptions based upon personal experience as the positive correlation of neuro psychic hypertension and glycosuria, dental caries and appendicitis which have been rather discredited by his own analyses of naval morbidity statistics; one feels quite indifferent to the result, if any, of the present compilation, being determined to perpetrate the paper anyway to stimulate possibly some other doubter to a more profitable investigation.

The use of statistics inductively to prove a theory would not be unique, but it is herein conscientiously denied.

Unfortunately one must make pretension to close familiarity

Digitized by Google

with naval morbidity data, their inherent and special defects as well as those common to the medical profession and vital statistics at large; for this reason one must decline to defend in detail, which would require many pages, his selection of diseases and methods of determining comparative loss rates.

Pursuant to the foregoing preamble the incidence, invaliding and death rates for 5 classes of scleroses have been plotted (Plate I) for the Navy and Marine Corps for the years 1871 to 1914 inclusive. Although 15 curves upon the same sheet may promote confusion it is believed that their composite represents more intelligently the movement of these diseases than would the resultants obtained by their combination; particularly will widely divergent accidental movements be declared for explanation without reflecting discredit upon steady tendencies.

Most of the titles selected for tabulation for 45 years have been arbitrarily grouped as follows; the only mathematical device utilized, and this really of negligible consequence, is the assumption during certain years that any patient deserting or going on leave represented half the loss of an actual invaliding from the service and has been computed as such:

(1) CIRCULATORY SCLEROSES.

Aneurysm	Intercerebral hemorrhage
Angina pectoris	Valvular disease, heart
Apoplexy	Varix
Arterial sclerosis	Myocarditis
Acute cardiac dilatation	Rupture of heart
Chronic cardiac dilatation	Cardiac hypertrophy
Embolism	Cordis inhibitio
Acute endocarditis	Cordis adiposum
Chronic endocarditis	Degeneratis cordis
Epidural hemorrhage	Atheroma

Paralysis cordis

(2) MALIGNANT PROLIFERATIVE SCLEROSES.

Carcinoma	Epithelioma
Sarcoma	Malignant hypertrophy of prostate

(3) PULMONARY SCLEROSES.

Asthma	Emphysema

Chronic bronchitis

Digitized by Google

(4) VISCERAL SCLEROSES.

Diabetes insipidus	Uraemia
Diabetes mellitus	Haematuria (non specific)
Gout, acute and chronic	Ascites
Gastric ulcer	Haematomesis
Prostatitis (non infective)	Albuminuria
Nephritis, acute and chronic	Hydrops
Lithaemia	Anasarca

Bright's disease

(5) NEURO-SCLEROSES.

Dementia	Monoplegia
Dementia paralytica	Paralysis agitans
Dementia precox	Sclerosis lateral
Hemiplegia (non traumatic)	Sclerosis posterior
Locomotor ataxia	Multiple sclerosis
Paraplegia	Acute ascending paralysis
Sclerosis disseminated	Cirrhosis cerebri
Amyotrophic lateral sclerosis	Senility

Syringomyelia

The chart conveys the first impression (a) that there was a sharp rise in nearly all rates in 1879; (b) that incidence rates have since tended to fall consistently except for the circulatory scleroses (organic cardiac and valvular disease) during the last 20 years; (c) that during certain years there have been wide fluctations in pulmonary and circulatory diseases; (d) that mortality rates have not fluctuated as widely as have incidence or even invaliding rates; (e) that during the last 20 years upwards of two thirds of the cases of circulatory and neuro scleroses have been invalided so that case invalidity and case mortality have risen decidedly which either is no credit to the clinician or reflects upon the diagnostics of earlier times. The opinion is ventured that the coefficient of unreliability of naval vital statistics would probably be proportional to the remoteness of decades with the maximum error prior to 1879 when decided improvement in collection and compilation occured; invalidings were first classified in published reports in 1879.

Any high rate for 1903 "is undoubtedly due to the large number of recruits admitted to the service during the year" according to the report of the Surgeon General covering that

Digitized by Google

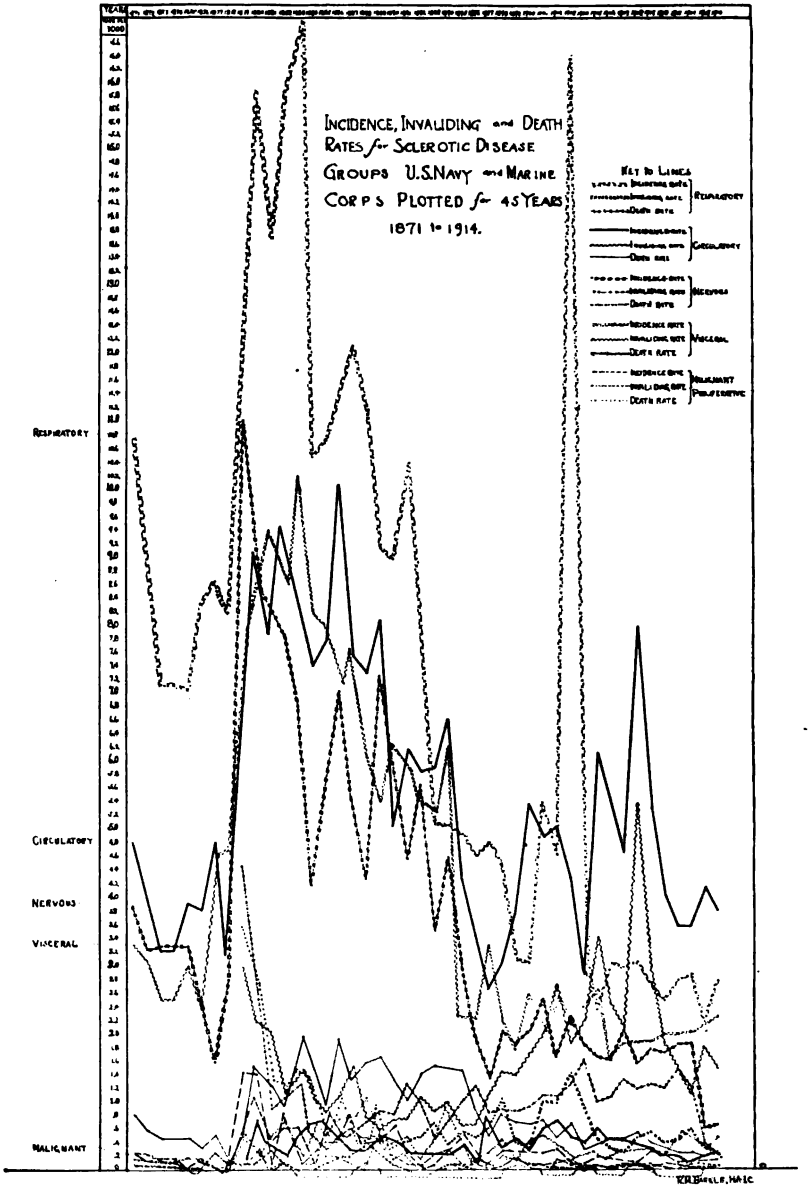


PLATE I.

Digitized by Google

year. There were accepted 15 per cent. more enlisted men in 1903 than during 1902 and with a lower percentage of rejections than formerly on account of necessarily rapid work of travelling recruiting parties which enrolled nearly one half of the men, practically all recruits, and accepted 59.5 per cent. as contrasted with 53.7 per cent. for the previous year. There were 556 cases of "chronic bronchitis" during 1903 which, as compared with 103 the year before and 86 the year following, accounts for the high rate; this was a diagnosis frequently made in recruits with chronic cough whom it was not policy to retain long enough in the service to determine the underlying cause.

The rise in incidence and invaliding rates for circulatory scleroses in 1905 and 1908 is similarly explained being largely accounted for by recruits with "chronic valvular heart disease," who should never have been accepted. There is, nevertheless, in spite of explanations of variable recruiting influence, a perceptible resultant tendency for circulatory rates to rise and this closely parallels the curve for admissions for syphilis which gradually rose to 27.18 per thousand in 1911 and have since declined to 19.84; for present purposes one is more concerned with the sequelæ of venereal diseases than their incidence and the fact that the invaliding rate for all venereal disease has declined steadily from 4.28 per thousand in 1909 to 1.71 in 1914 may be assumed to account for the current decline in rates for circulatory scleroses.

Subject to the conditions indicated in the first part of this paper and to the qualification that small numbers produce rather sharp fluctuations, one concludes that degenerative diseases in the United States Navy and Marine Corps have decreased during the past four decades and to the extent that these data may be considered to reflect the movement of cancer, paralysis, nephritis, and chronic heart, vascular, and pulmonary diseases (except tuberculosis) among the general population there is sufficient indication of similar decreases to justify withholding the prevalent assumption that these diseases are tending to increase until more scientific precision can be obtained in general morbidity statistics.

Digitized by Google

FLORENCE NIGHTINGALE AS STATISTICIAN.

By Edwin W. Kopf.

The somewhat legendary accounts of this remarkable woman contain but few references to that part of her life and work which should appeal to the students of the history of modern social statistics. More or less is understood of her radical innovations in the nursing care of the sick in institutions and especially in military hospitals; a definite idea exists of her capacities as reformer, administrator, and nurse. Comparatively little is known of her, however, as a constructive compiler and interpreter of descriptive social statistics. One biographer alone, Sir Edward T. Cook, speaks with sympathy and understanding of her as a statistician. He calls her a "passionate statistician."*

The activities of Miss Nightingale in statistics may be classed under several broad categories. We may think of her in terms of her forty years of thought and achievement in the Indian question; in safeguarding the health of the British soldier; in reorganizing civil and military hospital administration at home and abroad; and, in this latter regard, of her pioneer services to the profession of nursing. Her keen intellect, applied to these major projects of her career, comprehended the utility of the statistical method as a means of developing a basis of established fact for social reform.

In early life, Miss Nightingale showed peculiar aptitude for collecting and methodically recording current historical facts. Her observations during the travels of the Nightingales in Europe over the period 1837–1839 are a curious mixture of comment and criticism on the then existing laws, land systems, social conditions, and benevolent institutions. Throughout her life she collected an immense number of pamphlets, reports, and returns which she skillfully analyzed with telling effect in her campaigns for hospital and sanitary reform. Following her nursing apprenticeship with the Fliedners at

* The writer is indebted to Sir Edward T. Cook's "Life of Florence Nightingale" for abstracts from Miss Nightingale's private papers, and for much of the material in this paper.

Digitized by Google

Kaiserswerth, she undertook further training at the *Maison de la Providence* in Paris; here she proceeded to collect hospital reports, returns, statistical forms, and general information on hospital construction and sanitation. Among her papers then were elaborately tabulated analyses of hospital organizations and nursing systems and their end results. These inquiries extended to both France and Germany. She seems also to have addressed circulars of inquiry on the same subjects to representative hospitals in the United Kingdom.

Miss Nightingale was profoundly influenced by the works of Adolphe Quetelet, the Belgian astronomer, meteorologist, and statistician. Perhaps her practice of methodically recording the facts of her botanical researches led her to one of Quetelet's laws of flowering plants. The common lilac flowers, he averred, when the sum of the squares of the mean daily temperatures, counted from the end of the frosts, equals 4264° centigrade. While this "law" delighted her, she regarded it as a lesser example of Quetelet's researches and statistical conclusions. She was fascinated most by Quetelet's "*Sur l'Homme et le Développement de ses Facultés*," published in 1835, in which he outlined his conception of statistical method as applied to the life of man. From Quetelet, Miss Nightingale learned much of the science and art which describes human society in terms of numbers. From him she learned the methods, general aims, and results of qualified inquiry into social facts and forces.

MILITARY AND SANITARY STATISTICS OF THE CRIMEAN WAR.

The discipline of Quetelet's new science of social inquiry was to have its first influence upon the military and sanitary statistics of the Crimean War. Miss Nightingale found the medical records of the Scutari hospitals in lamentable condition. Even the number of deaths was not accurately recorded. The three separate registers then maintained gave each a totally different account of the deaths among the military forces. None of the statistical records was kept in uniform manner. She was able to introduce an orderly plan of recording the principal sickness and mortality data of the military hospital establishments which came within the sphere of her influence.

Digitized by Google

Miss Nightingale's experience in the Crimea filled her with an ardent desire to remedy the scandalous neglect of sanitary precautions in the Army; her study of the available data convinced her that the greater number of deaths in hospitals need not have occurred at all. During the first seven months of the Crimean campaign, a mortality of 60 per cent. per annum from disease alone occurred, a rate of mortality which exceeded even that of the Great Plague in London, and a higher ratio than the case mortality of cholera. Miss Nightingale's vigorous use of these facts resulted in a series of reforms, which in turn reduced this terrible rate of mortality. She observed, also, that if sanitary neglect prevailed in the Army afield, it probably affected the Army at home in considerable degree likewise. She compared the mortality in civil life with the mortality in army barracks. Between the ages of 25 and 35 she found that the mortality among soldiers was nearly double that in civil life. In writing to Sir John McNeill she said: "it is as criminal to have a mortality of 17, 19, and 20 per thousand in the Line, Artillery and Guards, when that in civil life is only 11 per 1,000, as it would be to take 1,100 men out upon Salisbury Plain and shoot them."

Her further observations on the Chatham military hospitals were: "This disgraceful state of our Chatham Hospitals is only one more symptom of a system, which, in the Crimea, put to death 16,000 men—the finest experiment modern history has seen upon a large scale, viz., as to what given number may be put to death at will by the sole agency of bad food and bad air." Among her private notes of 1856 her biographer found this: "I stand at the altar of the murdered men, and while I live, I fight their cause." The one weapon upon which she placed most dependence was her collection of sanitary statistics.

HEALTH, EFFICIENCY, AND HOSPITAL ADMINISTRATION OF THE BRITISH ARMY.

The results of her personal studies of army medical statistics were embodied in a report, from the first intended as a confidential communication to the War Office and the Army Medical Department. There had been considerable delay in

Digitized by Google

the formation of the Royal Commission on the health of the Army which she had requested in her November, 1856, interview with Lord Panmure. The Royal Warrant establishing the Commission was not issued until May 5, 1857. During this exasperating period of delay Miss Nightingale held in reserve her array of statistics, until, having begun her agitation with the sovereign and continuing through the politicians, she was almost ready to plead her cause with the people. In three months from the day the Royal Warrant was issued, the Commission presented its report. In the meantime, Lord Panmure had asked Miss Nightingale for her "Notes Affecting the Health, Efficiency and Hospital Administration of the British Army." These notes are the least known of her works, because they were never officially published. It has never become known how much of the final Report of the Royal Commission was actually the work of Miss Nightingale. Printed at her private expense and circulated among influential people, her "Notes" made a profound impression. They have been termed "a treasury of authentic fact . . . affording a complete elucidation of the causes which had brought about failure, and showing the means by which the country could best hope to safeguard the truly sacred task of providing for the health of its troops in future wars." Another of her friends who read the proof said: "It has so much the character of good, sincere, enlightened conversation on a subject which is thoroughly understood and appreciated, and so little the appearance of having been 'got up' or of pretension of any kind, literary or artistic." Another reader said: "I regard it as a gift to the Army, and to the country altogether priceless."

The preface to the Notes gave the keynote. Hospitals were shown to be but part of wider programs involving the general health and efficiency of the Army. This was emphasized by the fact that those who fell before Sebastopol by disease were above seven times the number who fell by the enemy. The introductory chapter gave the history of the health of the British Armies in previous campaigns. Six of the twenty sections of the Notes dealt with the medical history of the Crimean War. Two other sections discussed the mortality

Digitized by Google

of armies in peace and war and the necessity for a statistical department of the army. There were also numerous appendices, supplementary notes, and graphic illustrations and diagrams.

PIONEER IN GRAPHIC ILLUSTRATION OF STATISTICS.

It must be remembered that these Notes were written by Miss Nightingale in the short space of six months, and while in delicate health. In the preparation of her report she had but little assistance; the gathering of the data was facilitated, however, by the friendly coöperation of many broadminded men in the public service. Dr. Farr, for instance, aided materially in the preparation of the comparisons between the mortality of civil and army life and in editing the graphical illustrations which he especially commended. These graphical diagrams were at that time somewhat of an innovation in statistics, and had no significant precedent save in the statistical works of A. M. Guerry, a contemporary of Quetelet. The Report of the Commission, containing some thirty-three written answers by Miss Nightingale to leading questions by the Commission, together with her original tabulation of the appalling morbidity and mortality statistics of the British Army, was issued to the public in January, 1858. The graphical illustrations in her own Notes portrayed, by means of shaded or colored squares, circles and wedges, (1) the deaths due to preventable causes in the hospitals during the Crimean War and (2) the rate of mortality in the British Army at home. "Our soldiers enlist," as she put it, "to death in the barracks." She reprinted this graphic section and distributed it, with a brief memorandum, to leading members of Parliament and to medical and commanding officers throughout the country, in India, and in the Colonies. "It is our flank march upon the enemy," she said.

STATISTICAL DEPARTMENT OF BRITISH ARMY FOUNDED.

The chief product of the Commission's work of interest from the statistical standpoint was the report of the subcommittee on Army Medical Statistics. This committee, consisting of Mr. Sidney Herbert, Sir A. Tulloch, and Dr. Farr, reported

Digitized by Google

in June, 1858, and published its "First Annual Statistical Report on the Health of the Army" in March, 1861. The compilations were directed by Dr. Thomas Graham Balfour, under whose leadership British army statistics became the best and most useful obtainable in Europe.

The facts published in the "Notes" did not go unchallenged. A pamphlet appeared anonymously calling them in question. Miss Nightingale immediately prepared a reply. This second note was entitled "A Contribution to the Sanitary History of the British Army during the late War with Russia," and constituted a scathing and eloquent account of the preventable mortality which she had witnessed in the East. The graphic charts of the "Notes" were reproduced.

CONSTRUCTION, ORGANIZATION, AND MANAGEMENT OF CIVIL HOSPITALS.

The opposition to the recommendations of the subcommission on Army Barracks stimulated Miss Nightingale to prepare a more extended discussion of hospital construction, organization, and management. From her extensive experience in and study of hospital systems in Germany, France, and Ireland and in the Crimea, she prepared two addresses on hospital construction and sanitation for the Liverpool meeting of the National Association for the Promotion of Social Science. These papers were reprinted as "Notes on Hospitals." These "Notes" in three editions, the last in 1863, revolutionized ideas of hospital construction.

It was pointed out that the hospital statistics then available gave little information of real value on the proportion of recoveries, of deaths, and the average duration of hospital treatment for different diseases, duly qualified by sex and age. A common agreement on the number and nature of statistical data to be tabulated was recommended. A unique feature of this Liverpool address was a mortality table for hospital nurses and attendants showing the greatly increased prevalence of communicable diseases among this class of hospital employees, as compared with the mortality from the same causes in civil life. The deplorable existence of "hospital gangrene" and "hospital septicemia" in that day of defective hospital

Digitized by Google

sanitation and construction was effectively portrayed by these mortality statistics.

A brief inquiry into the precedent circumstances will be of interest. When Miss Nightingale returned from the Crimea she directed much thought and attention to hospital statistics as an adjunct to administration of institutions for the care of the sick. She found a complete lack of scientific coördination. The statistics were not kept along uniform lines. Each hospital followed its own nomenclature and classification of diseases. The available data had never been tabulated upon forms which would render the statistics of one hospital comparable with those of another. The data had little value for advancing medical knowledge or as an adjunct to hospital management. With the assistance of Dr. Farr, and of other friendly physicians, she drew up a standard list of diseases (largely a selection from the d'Espine-Farr System) and a set of model hospital statistical forms. She had her model forms printed in 1859 and persuaded some of the London hospitals to adopt them experimentally. She and Dr. Farr studied the tabulated results, which had sufficient value to show how large a field of qualified statistical inquiry had been opened by the introduction of her forms.

Miss Nightingale's skill in so effectively employing the statistical method in army sanitary reform had led to her election, in 1858, to fellowship in the Royal Statistical Society. On October 16, 1874, the American Statistical Association elected her an honorary member. A photographic reproduction of her letter of acceptance appears on the following page.

INTERNATIONAL STATISTICAL CONGRESS OF 1860.

This growing association with the leaders of thought in the statistical world of her time enabled her to take an active part in drawing up the program of the second section of the International Statistical Congress, held at London in 1860. This Section dealt with sanitary statistics. Miss Nightingale and Dr. Farr incorporated the forms for uniform hospital statistics, which had been experimentally introduced into a group of London hospitals in 1859, in a paper read for her before the Section by Dr. McMillian. Additions and recommendations,

Digitized by Google

London Jan.y 30 1875

Sir

I request that you will be good enough & thank the American Statistical Association in my name for the honour which they have done me in electing me an Honorary Member & to assure them that I am ever their faithful servant

Florence Nightingale

I trust that I may be allowed to send a Blue Book, being the last "Annual Sanitary Report" of the India Office: & a pamphlet: "Life or Death in India" by myself: as a small contribution to their library:

F.N.

To the Corresponding Sec.y

Digitized by Google

Digitized by Google

chiefly by Dr. Berg of Sweden and Dr. Neumann of Berlin, were concurred in by the author in a letter to the Earl of Shaftesbury, President of the Section.

This paper on uniform classifications and forms for hospital statistics was afterward widely circulated among physicians and hospital officials. Large quantities of the forms were supplied to hospitals in various parts of the country. The Paris hospitals took up the plan. Guy's Hospital, London, prepared a statistical analysis of its experience for the years 1854 to 1861; St. Thomas' for the years 1857 to 1860; and St. Bartholomew's, for 1860. At a meeting held at Guy's Hospital on June 21, 1861, it was unanimously agreed to adopt a uniform plan of registration, that each hospital should publish its own statistics annually, and that the forms devised by Miss Nightingale should be used so far as practicable.

ELEMENTS OF HOSPITAL MEDICAL STATISTICS.

Miss Nightingale then prepared the detailed paper on "Hospital Statistics and Hospital Plans" for the Dublin meeting of the National Association for the Promotion of Social Science in 1861. In this paper she emphasized the seven primary tabulation elements of hospital sickness statistics, which were:

(1) Number of patients remaining in hospital on first day of year.

(2) Patients admitted during year.

(3) Patients recovered or relieved during the year.

(4) Patients discharged as incurable, unrelieved, for irregularities or at own request.

(5) Patients died during year.

(6) Patients remaining at end of year.

(7) Mean duration of cases in days and fractions of a day.

These tabulation "elements" were to be compiled as seven separate tables, each showing diseases classified thereunder by sex and age (by single years under five and by five year periods thereafter). The additional consideration of diseases contracted in hospital while under treatment was also provided for. These extensions of the original paper were suggested by Drs. Berg and Neumann. Miss Nightingale held that these supplementary tables would bring out the fact of the then

Digitized by Google

scandalous prevalence of "hospital diseases" such as gangrene and septicemia.

The statistics of the various hospitals adopting her forms were published in the *Journal* of the Royal Statistical Society for September, 1862. Miss Nightingale's system of uniform hospital statistics was never generally successful over any considerable period of time. The plan, but partly realized in the requirements of the King's Hospital Fund, demands for its complete and effective operation a more intelligent appreciation of and a finer enthusiasm for statistical facts than is afforded even by present day voluntary and competitive hospital systems in metropolitan districts.

A further example of Miss Nightingale's use of the statistical method in hospital economy was her study of the questions relating to the possible removal of St. Thomas' Hospital at the instance of the Southeastern Railway. The railway company proposed the removal of the hospital to provide for an extension of the right-of-way from London Bridge to Charing Cross. She analysed the origins of cases served by the hospital, tabulated the proportions of cases within certain radial distances, and showed the probable effect upon patients of the removal of the hospital to the several possible sites suggested. This method of fitting hospital accommodation to the needs of populations has only recently been revived. It represents a legitimate application of demographic principles to the study of the relief of dire human needs.

The statistics of surgical operations from the standpoint of hospital cost and practical end results were next considered by Miss Nightingale. In a commentary on St. Bartholomew's statistical report, Miss Nightingale outlined the minimum requirements of a report form for the nature and result of surgical operations. The subject was further developed in a paper read for her before the Berlin meeting of the International Statistical Congress in 1863.

Before the close of the London meeting of the International Statistical Congress, 1860, Miss Nightingale addressed a letter to Lord Shaftesbury. The letter was read to the whole Congress and adopted by it as a resolution. The resolution impressed upon governments the prime necessity for publishing

Digitized by Google

more extensive and numerous abstracts of the statistical information in their possession.

MISS NIGHTINGALE AND THE CENSUS OF ENGLAND, 1861.

Miss Nightingale made a determined effort to extend the scope and application of the Census of 1861, largely in the direction of collecting statistics which would serve as a foundation for sanitary reform. Her aim was twofold: one was to enumerate the sick and infirm on Census Day. To those who denied that it could be accomplished, she pointed out that it had been done elsewhere, notably in Ireland. Her second ambition was to obtain complete data on the housing of the population; this, too, had been practicable in Ireland, she urged. In pressing her point with the Census officials she said: "The connection between the *health* and the *dwellings* of the population is one of the most important that exists. The 'diseases' can be approximated also. In all the more important—such as smallpox, fevers, measles, heart disease, etc. —all those which affect the *national* health, there will be very little error. Where there *is* error in these things, the error is uniform . . . and corrects itself. . . . " These few remarks still serve as prolegomena to any future census plans for England or any other country.

The Census Bill came up late in the session and not much comment can be found on the foregoing suggestion for a development of Census inquiry. The only critical comment made in the debate proceeded from Lord Ellenborough, who, far from considering the innovations of sickness and housing statistics, proposed to exclude most of the inquiries already suggested. Miss Nightingale, in her conception of census methods and results, was far ahead of her day and generation. Subsequent censuses, chiefly in the United States, in Tasmania, and in one or two other countries, have included sickness and housing inquiries.

SCOPE AND USE OF STATISTICS.

Miss Nightingale's activities in furthering statistical progress were the outgrowth of her deep conviction, variously expressed in her several papers, that the social and moral

Digitized by Google

sciences are in method and substance statistical sciences. In her several papers on metaphysical topics, she asserts that statistics were to her almost a religious exercise. Her conception of theology was that its true function was to ascertain the "character of God." Statistics, she mused, discovered and codified law in the social sphere and thereby revealed certain aspects of "the character of God." Doubtless, in these speculations she was profoundly influenced by the studies of Quetelet in moral statistics, as typified in his "*Recherches statistiques sur le Royaume de Pays-Bas.*"

Statistics as an element in political education also appealed with peculiar force to Miss Nightingale. In a letter to Benjamin Jowett she said: "The Cabinet Ministers, the army of their subordinates . . . have for the most part received a university education, but no education in statistical method. We legislate without knowing what we are doing. The War Office has some of the finest statistics in the world. What comes of them? Little or nothing. Why? Because the Heads do not know how to make anything of them. Our Indian statistics are really better than those of England. Of these no use is made in administration. What we want is not so much (or at least not at present) an accumulation of facts, as to teach men who are to govern the country the use of statistical facts."

She proposed a number of leading questions which she desired to see investigated by the statistical method: What had been the result of twenty years of compulsory education? What is the effect of town life on offspring in number and in health? What are the contributions of the several social classes to the population of the next generation? In proposing these inquiries she anticipated by more than a generation the work of the eugenists and biometricians of the Galton-Pearson school. Her friend, Adolphe Quetelet, had inaugurated studies of this character. Both he and Dr. Farr had hoped that she would pursue her inquiries more extensively. In conversation with Benjamin Jowett she proposed to found at Oxford a Professorship or Lectureship in Applied Statistics. Mr. Jowett seems to have discussed the matter with Mr. Arthur Balfour and Professor Alfred Marshall. Miss Night-

Digitized by Google

ingale, on her part, consulted Mr. Francis Galton, who responded earnestly and worked out a detailed plan. There was more or less discussion over the matter, but in the press of other affairs, the proposition was discarded. It is certainly of very great interest in this connection to observe that many years afterward such an enterprise was undertaken by the University College in London; probably, however, not directly issuing from Miss Nightingale's suggestion.

VITALITY OF ABORIGINAL RACES.

Following her disquisitions on army, medical, and civil hospital statistics, came her statistical investigations into colonial questions: the first into the vitality of native or aboriginal races in the Colonies; the second into the sanitary condition and material welfare of the population and military establishment of India. In a paper before the National Association for the Promotion of Social Science, meeting at Edinburgh in 1863, she discussed the gradual disappearance of the native races when brought into contact with the influences of civilization. The paper was suggested by Sir George Grey, who, at that time, was deeply engaged with questions of Colonial policy. The preliminary inquiry related to the probable effect of European school usages and school education generally, upon the health of children, of parents, and of races which had not heretofore been brought under any system of education. With the assistance of the Duke of Newcastle, she prepared a form which was sent by the Colonial office to the governors of the various colonies. From the returns of 143 schools she deduced the mortality of school children by age period and by sex, and further classified the statistics by causes of death. A second inquiry into the statistics of colonial hospitals gave important information on the causes of high institutional mortality among the native races. The numbers involved were small, and the results were necessarily considerably in error; but, in the main, the conclusions as to the neglect of sanitary precautions and the change in the living habits of native races were sound.

A further paper on the aboriginal races in Australia was read before the same Association at York, England, in 1864. This

Digitized by Google

essay contained copious quotations from correspondence with colonial governors over points raised by the first paper. The Colonial Office circulated the reprint of the Edinburgh paper widely. Miss Nightingale has been considered a pioneer in work for arresting the decline of native races, so far as such work has been possible.

BRITISH ARMY IN INDIA.

After the death of Sidney Herbert, Miss Nightingale devoted the larger portion of her time and attention to the Indian question. Her earlier years of service to the British Army at home were not more significant than her later years of endeavor for the army and people in India. The greater proportion of the lives of soldiers lost in India was not chargeable to battle, but to disease caused by insanitation and general ignorance of military tropical hygiene. In 1859, it was found that the average annual death rate of the British armies in India had been 69 per 1,000 since the year 1817.* In recent times the figure has been 5 per 1,000. The changes in living and working accommodations which brought about this reduction in army mortality are directly traceable to the recommendations of the Royal Commission, appointed in 1859, which reported in 1863. An unrecorded fact, however, is that Miss Nightingale's suggestion led directly to the appointment of the Commission and that the greater part of the Report was her handiwork. The suggestions upon which permanent reforms in army sanitation in India were based were also her work.†

For eight months, during the latter part of 1858 and the earlier part of 1859, Miss Nightingale importuned Lord Stanley for the appointment of a Royal Sanitary Commission which would do for the armies in India what had been done for the armies at home. She had contemplated for two years before,

* While this figure included battle casualties of the campaigns of the forty years ending with 1856, the mortality figure was still excessive. It was stated that the registered mortality among British troops in India was six times that of Englishmen of the same ages at home. Again, in an earlier investigation into mortality from disease among troops in the East Indies (including for the most part British India) death rates varying from 40 to 98 per 1,000 of mean strength over the period 1840–1848 were found.

† British soldiers in India today live in barracks which surpass in comfort and sanitation any that can be found in other countries. Every regiment, battery, and depot has its regimental institute, a sort of soldiers' club, library, reading and recreation room, a temperance association, and a theatre. The use of alcoholic beverages is discouraged and every encouragement is given to useful employment for the men.

Digitized by Google

the appointment of a Commission to investigate the entire question of the armies in British India. The mutiny of almost the whole of the Bengal native army and the contingents in northern India in 1857, which had filled the minds of the British population with thoughts of vengeance and repression against the native army, had only served to fix her attention upon sanitary and other administrative reform on behalf of the soldiers. Her analysis of the statistics of army mortality in India convinced her that there was murder committed not by the Sepoys alone. To her mind, it was murder to doom British soldiers to death by neglect of the most elementary sanitary precautions.

Anticipating the appointment of the Commission, she began collecting, tabulating, and interpreting data she derived from circulars of inquiry which she had drafted and sent to all the stations in India. The inquiry form lacked little in requisite completeness and precision of detail. In the meantime, Miss Nightingale and Dr. Farr searched the sickness and mortality records of the India Office.

The report of the Indian Sanitary Commission when issued in 1863 comprised in all 2,028 pages, mostly in small print. The greater part of the statistical work in the Report bore clear evidence of Miss Nightingale's influence. Her inquiry blank, in the first instance, provided the vehicle for the transmission of much of the evidence. The replies to her questions occupy the whole of the second volume of the Report. In October, 1861, the Commission requested her to submit her interpretation of these Stational Reports. Her observations upon these reports occupied 23 pages and are the most remarkable of her published works. Her unusual treatment of the subject by the addition of illustrative wood-cuts describing Indian hospitals and barracks, made the Treasury demur, however, at the cost, but Miss Nightingale was permitted to pay for the printing out of her private purse. Copies of these observations were sent to the queen and to influential members of the government. Sir John McNeill wrote: "The picture is terrible, but it is all true. There is no one statement from beginning to end that I feel disposed to question and there are many which my own observation and experience enable me to con-

Digitized by Google

firm." In detail, her notes related to the camp diseases which follow the selection of poor sites, defective disposal of human wastes, overcrowding in barracks, lack of suitable occupation and exercise, dietaries and defective hospital arrangements. The sources of statistical data for the armies in India were also criticised. In addition to her observations, Miss Nightingale prepared with Dr. Sutherland an abstract of the returns upon which her "Observations" were founded.

Moreover, when the Report was published she moved with her characteristic decision to secure for it the newspaper and periodical publicity which it deserved. She contributed a popular résumé of the Royal Commission's Report to the National Association for the Promotion of Social Science, meeting at Edinburgh in 1863, entitled "How People may Live and not Die in India." The paper was republished in 1864, with a preface and an account of what the Commission had actually achieved in sanitary works and measures. In 1868, the Secretary of State printed her résumé of the Indian Sanitary Question from 1859 to 1867 and her memorandum of advice and suggestions on the entire situation as it then stood. The dispatch from Sir Stafford Northcote under date of April 23, 1863 (drafted by Miss Nightingale), was printed at the same time. This dispatch resulted in the first of the annual series of Indian Sanitary Reports. In the reports for 1868 and 1869 appear two of Miss Nightingale's contributions; in the first an Introduction of eight pages and in the second her paper on "Sanitary Progress in India."

Her statistical enterprises of this period are well summarized in a few phrases which are here quoted from one of her letters: "I am all in the arithmetical line now . . . I find that every year . . . there are in the Home Army, 729 men alive every year who would have been dead but for Sidney Herbert's measures, and 5,184 men always on active duty who would have been 'constantly sick in bed.' In India the difference is still more striking. Taken on the last two years, the death rate of Bombay is lower than that of London,* the healthiest city in Europe. And the death rate of Calcutta is lower than that of Liverpool and Manchester. But this is not

*These figures related only to conditions prevailing at the time this letter was written.

Digitized by Google

the greatest victory. The Municipal Commissioner of Bombay writes that the 'huddled native masses clamorously invoke the aid of the Health Department if but one death from cholera occurs; whereas formerly half of them might be swept away and the other half think it all right.'"

In 1873, the National Association for the Promotion of Social Science invited her to contribute a paper on the ten years of progress in India since her "How People may Live and not Die in India" paper appeared. Her title for this later paper was "How some People have Lived and Not Died in India." The India Office reprinted the paper in its 1874 bluebook. The salient fact developed in this report was that the death rate in the Indian Army had been reduced from 69 per 1,000 to 18 per 1,000. This summary of ten years of sanitary progress in India she was qualified to prepare in consequence of her editorial work on the annuals issued by the sanitary department of the India Office.

In 1877, Miss Nightingale published two letters on famine in India and followed these by an article in the *Nineteenth Century* magazine. This article, "The People in India," gave the principal facts about the Indian famines and proceeded further to describe in considerable detail the evils of usury in the Bombay Deccan. Beginning with 1874, Miss Nightingale collected statistics of irrigation in India, and of the effect of irrigation on the life and health of the people. These data, the appendix of the second part of an unpublished work on the Indian Land Question and Irrigation Systems, were afterward partially used in several isolated papers. She thought much on education in India. There had been a neglect of elementary education. The exception was found in the system of village schools established by Lieutenant-Governor James Thomason, of what is now the Agra Province. The report of the Indian Education Commission of 1883 directed attention to the essential difficulties residing in the language, credal, race, and traditional differences of the populations of the several provinces. The two chief difficulties in the way of a diffusion of education among the masses in India are the large agricultural population, among whom it is in all countries difficult to advance any system of education, and the existence

Digitized by Google

of a hereditary class, whose object has been to maintain their monoply of learning as the chief buttress of their social supremacy. Questions such as these occupied Miss Nightingale's attention in the years 1881 and 1882.

The succeeding years were taken up in turn by Army Hospital Service reform, district nursing organization, nursing education and Indian finance problems.

In 1891 Miss Nightingale finally laid aside her ambition to found at Oxford a Professorship in Applied Statistics. This relinquishment of active interest in the progress of statistics was indicative only of her failing physical powers. The gradual failure of sight, memory, and mental apprehension proceeded during the last fifteen years of her life. Her death occurred on August 13, 1910.

FLORENCE NIGHTINGALE AS STATISTICIAN.

Florence Nightingale may well be assigned a position in the history of social statistics next to those occupied by Quetelet and Farr. Her ardent, genuine sympathy for the sick and distressed was greatly augmented by a positive genius for marshalling definite knowledge of the forces which make for disease and suffering. The same intellect which sharply separated the formulae, procedures, and practical methods of nursing from its abiding principles as one of the humanities, also discerned the statistical facts of sickness and other forms of disharmony between the individual and his environment. In hospital care of the sick, as an instance, Miss Nightingale replaced the astigmatic case viewpoint with one embodying a grasp of total situations. This is one function of statistics.

Her earnest perception of this truth—an essential in the equipment of the statistician—was firmly supported by her control over laborious detail and by her scrupulous care in testing the statistical foundations of her premises. The interpretations she placed upon the facts developed in her researches show a careful regard for the competent counsel which she so often consulted. In all these respects Miss Nightingale exhibited the prime qualities of one thoroughly versed in the art of preparing and reflectively analysing social data.

Digitized by Google

CONSUMPTION OF GAS AS AFFECTED BY POPULATION.*

By G. P. WATKINS and L. H. LUBARSKY.

The relation between population and the amount of use of transportation facilities has often been made a subject of statistical investigation. As regards street railways, in particular, it has been claimed that the number of passengers varies as the square of the population of the cities served. This hypothesis, as thus mathematically expressed, is questionable, but there can be no doubt that there is a rough functional relation between the population of a city or cities and the number of street railway rides, and that the increase of the latter is much more than in direct proportion to the increase of the former. If the choice is between mere proportional variation and the quadratic relation, the latter is much nearer the truth.†

The possibility of a similar relation between agglomeration (and, in a somewhat loose sense, density) of population and the extent to which the services of other public service enterprises are availed of is worthy of investigation, though the others are not so easily dealt with as the street railways. In the case of water supply the method of distributing the cost accords more with principles of taxation than with those of market price-making. The competition between electricity and gas—as regards use and consumption, if not always commercially—which is facilitated by the overlapping of the means of service, causes difficulties in the case of these utilities. It would be desirable to deal with light, heat, and power as separate services rendered the consumer, instead of with gas supply and electricity supply, thus avoiding the complication due to the substitution of electricity for gas. The present study is subject to an important limitation on this account.

*The general division line between the work of the joint authors of this paper is: economic and statistical analysis mainly by Watkins, and graphic construction and analysis mainly by Lubarsky.

†For a discussion of this question by the writer of the above paragraph, see the Annual Report of the New York Public Service Commission for the First District for 1910, Vol. III, pp. 22–53; also a qualification of the conclusion in the similar report for 1912, Vol. II, p. 109, including the footnote.

Digitized by Google

In all these cases, there is the dynamic as well as the static aspect of the situation to be considered. The latter viewpoint relates to two or more cities of different size compared for the same date; the former relates to the same city at different times with regard to intervening growth of population. Time and progress enter greatly into the dynamic kind of comparison and but little into the static. The interest of the sort of study in question is at least as much dynamic as static. But the way to obtain an indication of the importance of the former element is to determine and make allowance for the importance of the latter, since even where time brings in other and possibly more important influences, mere population magnitude is a factor. Hence the two aspects of the general problem need to be studied together. It is too often assumed that there is a constant relation between population and consumption, or that the only cause of change in per capita consumption is a change in price. This assumption does not hold for urban public utilities, though it may sometimes be well enough, pragmatically speaking, to assume constancy of consumption, apart from price changes, for short periods of time.

In the case of gas supply, it would be quite easy to attribute to differences in population results that are actually due to price differences, since the price per thousand itself bears in a general way a functual relation to the size of cities. Companies serving larger cities can usually afford to supply gas at lower rates. The lower price is, of course, a reciprocal effect of density of demand (larger per capita consumption) and of consequent "increasing returns," so that differences of the kind in question need not be dismissed as simply due to differences in price. But allowance needs to be made for this factor, that is, for the stimulating effect of a lower price upon consumption. The nearly uniform nickel street railway fare from place to place as well as from time to time in the United States is an important statistical advantage that is wanting in the case of the study of other utilities.

A dynamic factor of some independent interest is an increasing standard of living. The city dweller has doubtless considerably raised his standard in the past generation, at least as regards degree of per capita use of public utilities. As

Digitized by Google

higher
gas as
ooking

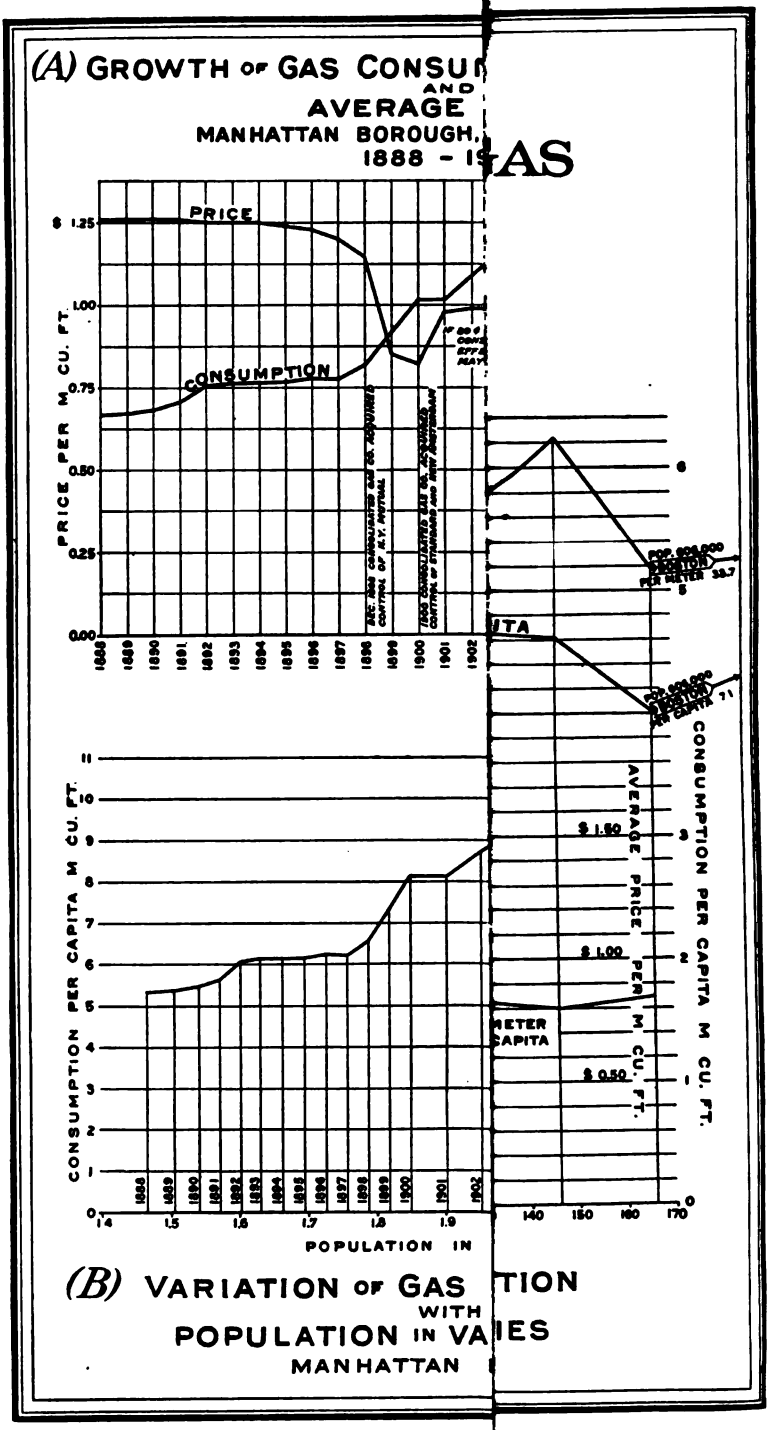
or per
in one
statis-
raphic
he ac-
grams,
capita

of the
ption
from
art in
erage
this
ction
gures
tions.
tory,
g on

e for
the
rent
re is
aries
reat
ities
mp-
im-
rre-
hile
of
the

figures

Digitized by Google



Digitized by Google

regards gas consumption, it should be noted that the higher standard often means the substitution of electricity for gas as an illuminant, and, on the other hand, of gas for coal in cooking and to some extent in other heat uses.

How much greater is gas consumption per consumer or per capita in a community having a large population than in one having a small population, is the subject of the present statistical investigation. The question is dealt with by the graphic method on the basis of such data as are available. The accompanying diagram which consists of three sub-diagrams, lettered A, B, and C, serves to determine how far per capita consumption is a function of varying population.

Sub-diagram A, the one in the upper left-hand quarter of the plate, shows primarily the growth in per capita consumption of gas in the Borough of Manhattan, City of New York, from 1888 to 1914. As the price of gas plays an important part in influencing consumption, another curve showing the average price per M cu. ft. for the same period is also shown in this sub-diagram. Sub-diagram B, in the lower left-hand section of the plate, utilizes the same data, but the consumption figures are geometrically coördinated with homologous populations. Both these sub-diagrams relate to the same urban territory, the variation in population and consumption depending on time.

The data for sub-diagram C have no relation to those for the other two. It shows that at the same point of time, the consumption of gas both per capita and per meter in different localities varies with the population of the locality. There is also an average price curve, which shows that price varies inversely as population; but the price variation is not as great as the variation of consumption. The data relate to localities in Massachusetts for 1910. The fact that, for gas consumption, figures of more recent date might be used is of little importance compared with the necessity of estimating corresponding populations that would accompany their use.* While serving mainly for comparison with and corroboration of results for New York City, the difference in the nature of the

*The data of the present article were put into shape and the diagram constructed before 1915 figures were available.

Digitized by Google

comparison should be noted. These Massachusetts figures are not directly affected by progressive changes with time—in other words, they show an economically static relation—while the New York City results may be due as much to the passage of time as to differences in population magnitudes—in other words, they are presumably dynamic in character.

The data used in the diagram are shown in the accompanying tables. The data of Table I are of independent interest. Those of Table II are of interest chiefly in connection with the diagram and therefore only such details are given as relate directly to the diagram and make it possible to identify the absolute numbers taken from the Massachusetts report. A detailed description of each sub-diagram together with a statement of its statistical significance follows the tables.

In connection with the population data of Table I, there is occasion to mention the probable deficiency of the 1890 federal enumeration of New York City with reference to its possible effect on the curves of the first two sub-diagrams. As is mentioned in a footnote to the table, allowance is made for this condition by arbitrarily increasing the Census figure for Manhattan Borough in 1890 by 100,000. According to the method employed to estimate population between decennial years the effect of any under or over estimation in the figure used is spread through the entire decade from 1890 to 1900 as well as before, and allowance could easily be made for it by the eye if occasion required.

As regards the Massachusetts localities, also, a qualification with reference to the population data is necessary. The population served by a particular gas company should strictly be, for comparative purposes, the number of inhabitants within reach of its mains. But population data of that nature are seldom to be had; nor does it make much difference if in the case of cities of medium or large size the territory taken in is determined by civil boundaries. In the case of small cities and towns some allowance for the effect of including people not within reach of the mains is necessary. Especially in Massachusetts, where the form of town government is such as to make the transition to organization as a city of little significance, the distinction between rural areas and areas

Digitized by Google

whose population is incorporated as villages and cities, is scarcely applicable, the town government being adequate for many urban functions and the city, when organized, being territorially coextensive with its predecessor town. In fact the United States Census classifies the population of Massachusetts towns as urban or rural merely according to the number of inhabitants a town contains, the significance of an "incorporated place" being applicable in the ordinary sense only outside of New England. Hence the population of the smaller Massachusetts localities used in connection with this table may be misrepresentatively large in proportion to the number of inhabitants in position to be served by the gas companies.

The data for sub-diagram A are the two columns of derivative numbers of Table I. A few historical facts, having possible bearing on the direction of the curves at particular times, are noted at appropriate places on the diagram. The construction of the two curves needs little comment. As regards the price curve from 1905 to 1908, some explanation is necessary. On May 1, 1906, the price of gas for private consumers in the greater part of New York City was fixed at 80 cents per M cu. ft. by an act of the legislature passed the winter before. The gas companies operating in Manhattan took the matter to the courts and, pending a decision, the companies were allowed to charge the consumers $1.00 per M cu. ft., but the 20 cents excess was to be returned to the consumer if the 80 Cent Gas Law was held to be constitutional. On January 4, 1909, the United States Supreme Court upheld the law in question and the gas companies proceeded to refund the excess charge. Under these circumstances, there is room for question as to just what price should be compared with consumption per capita, but the preference should doubtless be given to the higher price, with a qualification. Both averages are shown in the price curve, the full line representing the price actually paid. The abrupt drop at the beginning of 1909 shows the change in the rate actually paid. The dotted curve between 1905 and 1909 shows the average price if the law be considered effective on May 1, 1906.

The curve of consumption per capita shows an almost steady

Digitized by Google

upward tendency. The average consumption has practically doubled since 1888. The variations or increases from 1888 to 1897 were slight. From 1897 to 1903 there was a rapid increase of about 3 M cu. ft. This marked increase is explained by the great reduction in the average price of gas. There was a slight falling off in consumption per capita in 1908, reflecting the after effects of the 1907 panic. Since then the per capita consumption has been steadily rising, due largely to the final reduction in price to 80 cents.

The price curve shows a decrease of more than one third since the earliest year represented in the diagram. From 1898 to 1900 there appears to have been lively competition, which is reflected by the drop in the price from about $1.15 in 1898 to 82 cents in 1900. In 1901 the average price became about $1 per M and it remained there until the 80 Cent Gas Law became effective in 1909. But the effect of the reduction upon per capita consumption extends several years beyond this. From 1909 the price to the public has been constant, but per capita consumption has continued to move upwards. Throughout the period covered by the diagram the growth of the use of gas for industrial purposes and for cooking has doubtless been a factor in increasing per capita consumption.

Irregularities in the growth of population not reflected in the estimates used affect the detailed course of the curve considerably. It is probable that population increased less rapidly in 1913 and 1914 than in 1910-12. A slight peak appearing in 1907, also the salient angle at 1892 may be due to increases in population which the method of estimation of this element cannot well take account of.

Sub-diagram B is the graphic presentation of the same facts on a different plan. Of the coördinates of the series of points plotted, the abscissae represent particular populations, and the ordinates, corresponding per capita consumption. The scale for population is thus along the horizontal axis, and that for consumption along the vertical axis. Both axes have the same point of origin, which, however, is not shown on the diagram, since only the part of the space between the portion of the axes in question that contains the points plotted is significant. The population points for each year are located

Digitized by Google

on the horizontal axis, and at each of them ordinates are drawn which are numerically proportional to the per capita consumption in each year. The population of the intercensal years having been determined as an arithmetic series, the intervals on the axis of abscissae between census figures are equal. So far as known, the maximum population was reached in 1910, the 1915 State Census figures for Manhattan Borough being considerably less than those for 1910.* Hence the ordinates for 1911, 1912, 1913 and 1914 recede instead of advance, and are therefore distinguished as dotted lines. The curve connecting the extremities of the ordinates chronologically reflects the direct movement of population until 1910, and the retrograde movement from 1910 to 1914. This curve is the same as the consumption curve of the first sub-diagram, only the yearly intervals are variable instead of being constant, and the ordinates advance as long as population increases and recede when population decreases.

The steady increase in per capita consumption is seen as before. But it is further shown that from 1910 on, although the population decreased, per capita consumption continued to increase. Hence per capita consumption cannot be said to be simply a function of population magnitude. It is influenced by factors apart from those directly affecting the growth of population in the territory served. The business and amusement district of Greater New York would naturally be indirectly favored by population growth occurring outside the borough that contains the most important of such interests. The rising standard of living would be similarly effective. It should be noted that from the point of view of judging of the general direction of the per capita curve as related to the growth of population, the second sub-diagram serves the purpose better than the first, since the slope of the curve changes as the rate of increase of population changes, with the exception, of course, of such irregularities as cannot be recognized in interpolating between census years. The slope is more nearly uniform in this curve than in the one above, though the effect of price changes is still marked.

*The state enumeration of 1915 was probably deficient for New York City; nevertheless there appears to have been an actual decrease in the population of Manhattan Borough.

Digitized by Google

Sub-diagram C is based on Table II. The population figures are as of April 15, 1910, the meters as of June 30, 1910, and the metered gas as of the two year period from July 1, 1909, to June 30, 1911. By taking the average per year of metered gas for the two years, a figure is obtained that synchronizes with the date of the number of meters. The slight divergence of the population date is not important, especially for urban population.

The method of plotting the third sub-diagram is similar to that for the second. The horizontal axis denotes population and the vertical consumption per meter and per capita, also price per M cu. ft. The coördinates of each point plotted as a dot are the population and per meter consumption for a particular locality, and of those plotted as plus signs, population and per capita consumption. Hence there are a dot and a plus sign on every ordinate. To eliminate the irregularities resulting from the large number of small localities, the latter have been grouped as shown in part B of Table II, averages for population and for per meter and per capita consumption having been computed by groups. The average points for each of these various groups together with the points for the two large localities having populations of 165,900 and 146,000, respectively, are connected, giving the two curves of consumption per meter and consumption per capita, respectively. The largest locality is that served by the Boston Consolidated Gas, with 606,000 population, an average per meter consumption of 33.7, and an average per capita consumption of 7.1, the price being 80 cents per M. Owing to a limitation of the graphic method, it is impossible to locate these points on the diagram as drawn. The character of the curves as they would be extended beyond the right edge of the diagram is, however, indicated by arrows. The coördinates of the smaller localities are properly plotted individually, but the points are not connected.

As a necessary qualification of the significance of the other curves a price curve is added. The average price per M cu. ft. for each locality is given in the table, but only the average price for each group is plotted. In computing these average prices, sales to other gas companies have been eliminated.

Digitized by Google

The degree of certainty of the significance of the data would be greater if prices were constant. But with the aid of the price curve, allowance can be made for such differences. Differences in per capita consumption between large and small localities are doubtless partly due to the fact that it is in general the smaller communities that have the higher rates. But allowance for the price factor appears to leave a difference still to be accounted for by other factors, whose comparative importance will be considered presently.

There is much irregularity in the variation of per meter and per capita consumption with the size of population aggregates. But, as already indicated, the general trend is unquestionably towards increased consumption both per meter and per capita with increased population. For reasons already mentioned, less importance should be attached to the points and lines for localities under 20,000 population than to those for the larger localities. The small number of the cities in each group is also a reason why considerable irregularity above this population magnitude is to be expected.

The slope of the per capita curve in sub-diagram C is of course not directly comparable with that of the curve in sub-diagram B. According to the trends of these two curves as roughly determined from the two sub-diagrams an increase of population in the same locality (Manhattan) of about 13 per cent. is accompanied by an increase of 50 per cent. in per capita gas consumption, while a relative difference in population between two separate localities for the same date must be 500 per cent. of the smaller population in order to correlate with a difference of 50 per cent. in per capita gas consumption. Of course such a comparison must be very rough. And it must be qualified with reference to price differences. The 50 per cent. increase in per capita gas consumption in Manhattan is accompanied by a 36 per cent. decrease in price. The 50 per cent. difference in per capita gas consumption among the Massachusetts localities is accompanied by about the same 36 per cent. difference in price. If we could attribute the whole difference in consumption to differences in price, assuming also that these rough quantitative inferences are sufficiently correct, then the demand curve for gas might

Digitized by Google

be alleged to be a "constant outlay" curve, a one third decrease in price per unit being responded to by a one half increase in consumption, the aggregate price thus remaining the same. As to the relation between price and consumption the data would perhaps repay mathematical analysis by way of correlation coefficients.

From this result it might be inferred that price accounts for all the differences and changes in consumption. But for such a conclusion the comparison proves too much. It negates the influence of the dynamic as well as the static factors. The former certainly appear to count for something in the Manhattan curves. And as regards the static (Massachusetts) curves, while there is evidently a close inverse correlation between price and consumption, yet between 90,000 and 600,000 population, where there is the smallest range of variation in group average prices (10 cents) there is the greatest range in per capita consumption, that is, from 3.5 to 7.1.

Apart from the question of the influence of price, it is clear that dynamic factors are far more important than the mere size of the population aggregate. As a matter of fact, it is quite proper to class price among the dynamic factors, but it was the intention of the present study to abstract from the effect of price changes and price differences so far as possible. And there seems to be some ground for believing that mere aggregation of population does tend to increase gas consumption somewhat. The complexity of the problem is such that an attempt to specify how much is hardly warranted. And it may be necessary to conclude, at the same time, that the public appears to be able quickly to obtain for itself most of the benefit of density of consumption by means of a lower price, thus obscuring the effect upon consumption of agglomeration of population as such.

Digitized by Google

TABLE I.

POPULATION, SALES OF METERED GAS, AND REVENUE DERIVED THEREFROM, BOROUGH OF MANHATTAN, CITY OF NEW YORK, 1888–1914.

Year Ending Dec. 31.	Population. (a)	Sales of Gas (M cu. ft.) (b).	Sales per Capita. (M cu. ft.).	Revenue from Sale of Gas.	Average Price per M cu. ft.
1888	1,465,908	7,797,147	5.32	$9,824,405	$1.260
1889	1,503,562	8,075,509	5.37	10,175,141	1.260
1890	1,541,216	8,409,343	5.46	10,595,772	1.260
1891	1,572,104	8,819,804	5.61	11,122,953	1.261
1892	1,602,991	9,681,954	6.04	12,102,443	1.250
1893	1,633,879	9,979,014	6.11	12,473,767	1.250
1894	1,664,766	10,181,325	6.12	12,726,656	1.250
1895	1,695,654	10,402,830	6.13	12,899,509	1.240
1896	1,726,541	10,742,799	6.22	13,213,643	1.230
1897	1,757,429	10,900,965	6.20	13,081,158	1.200
1898	1,788,317	11,661,300	6.52	13,410,495	1.150
1899	1,819,205	13,325,943	7.33	11,327,052	.850
1900	1,850,093	15,036,787	8.13	12,330,165	.820
1901	1,900,660	15,476,533	8.14	15,167,003	.980
1902	1,951,227	17,004,702	8.71	16,834,655	.990
1903	2,001,794	18,725,618	9.35	18,538,362	.990
1904	2,052,361	19,415,818	9.46	19,221,659	.990
1905	2,102,928	19,540,243	9.29	19,540,243	1.000
1906	2,148,651	20,296,822	9.45	17,690,151 (c) 19,904,915	.872 .981
1907	2,194,374	21,227,252	9.67	17,003,439 (c) 21,094,699	.801 .904
1908	2,240,096	20,442,298	9.13	16,378,154 (c) 20,341,694	.801 .995
1909	2,285,819	21,656,335	9.47	17,287,177	.798
1910	2,331,542	22,583,176	9.69	18,037,196	.799
1911	2,293,846	22,730,971	9.91	18,154,275	.799
1912	2,256,150	23,247,816	10.30	18,571,457	.799
1913	2,218,454	23,015,070	10.37	18,385,758	.799
1914	2,180,758	22,940,883	10.52	18,322,321	.799
1915	(d) 2,143,061				

(a) Sources: United States Census for decennial years and New York State Census for 1905 and 1915, figures for intervening years being interpolated according to arithmetical progression. But the Census figure for 1890 is increased by 100,000 to allow for deficient enumeration.

(b) Sources: 1888–1905 City's Exhibit No. 46, in the printed record of the case of the *Public Service Commission* vs. *the Consolidated Gas Co.*, Vol. V, p. 5745, also reproduced in William D. Mark's "Finances of Gas and Electric Light and Power Enterprises," opposite p. 398; 1906–1914, company returns furnished to Public Service Commission and its predecessor, the Board of Gas and Electricity. The four gas companies serving Manhattan Borough are the Consolidated, Standard, New York Mutual, and New Amsterdam. For 1907 and 1908 some estimation is necessary in order to put the figures on a calendar year basis.

(c) Includes 20 cents per M cu. ft. paid by the public pending decision of appeal of 80 Cent Gas Law.

(d) The figure for 1915 is preliminary and 5,314 larger than the final reported figure for New York County, but the enumeration appears to be considerably deficient.

Digitized by Google

TABLE II.

POPULATION, SALES OF METERED GAS TO CONSUMERS AND NUMBER OF METERS, FOR CERTAIN LOCALITIES IN MASSACHUSETTS, 1910.(a)

(A) SIGNIFICANT DATA FOR EACH LOCALITY.

Population of Territory Served (thousands).	Number of Meters.	Average Price per M cu. ft. (b)	Average Consumption per Meter.	Average Consumption per Capita.
606.4	128,170	$ .80	33.68	7.119
165.9	26,560	.854	25.92	4.149
146.0	21,626	.806	31.21	4.624
119.3	20,540	.845	24.71	4.255
118.5	20,577	.927	27.23	4.728
112.1	23,413	.780	27.24	5.688
110.2	18,763	.905	22.81	3.885
107.7	21,771	.871	24.80	5.015
96.7	14,841	.857	26.01	3.994
90.9	12,542	.876	21.01	2.898
88.4	18,347	.979	20.04	4.157
88.0	12,517	.94	26.99	3.838
85.9	9,793	.867	24.57	2.801
59.4	8,620	1.11	17.55	2.546
56.9	10,107	1.11	19.99	3.552
44.1	7,815	.931	26.21	4.644
37.8	4,793	1.14	18.81	2.383
37.5	4,444	.987	25.57	3.031
35.7	5,616	1.14	21.31	3.353
32.6	3,501	1.43	14.76	1.448
28.4	2,349	1.19	12.87	1.066
28.4	5,642	1.36	18.08	3.598
27.4	1,048	1.00	13.26	0.507
26.0	2,135	1.65	16.23	1.331
25.4	3,202	1.28	15.82	1.994
24.8	2,316	1.04	21.91	2.047
24.4	2,721	1.10	23.26	2.594
23.2	2,990	1.00	19.37	2.493
21.3	2,135	1.50	12.96	1.297
19.4	2,725	1.24	20.68	2.900
19.0	1,728	1.48	15.57	1.418
18.7	2,726	1.17	18.24	2.666
16.2	2,210	1.20	23.05	3.141
15.3	711	1.40	21.26	0.988
15.2	903	1.40	22.12	1.310
15.1	1,231	1.59	13.34	1.086
14.9	1,950	1.39	16.33	2.131
14.7	541	2.09	14.38	0.529
13.1	968	1.53	19.88	1.472
13.0	1,104	1.14	14.20	1.203
12.9	1,208	1.51	18.50	1.726
12.9	354	1.79	13.12	0.360
12.1	856	1.79	14.27	1.006
11.6	1,119	1.63	11.25	1.090
11.3	1,197	1.30	28.69	3.038
9.9	1,092	1.56	12.31	1.363
9.4	594	1.62	12.88	0.813
8.8	530	1.31	18.23	1.101
8.5	384	1.55	23.63	1.063
8.5	863	1.61	13.83	1.400
8.0	840	1.57	17.26	1.809
6.7	740	1.54	12.95	1.422
5.8	230	1.92	15.03	0.598
5.6	160	1.29	13.91	0.394
5.1	98	1.72	12.10	0.232
4.9	140	2.15	6.70	0.191
3.7	361	1.15	19.96	1.943
3.0	365	2.07	9.35	1.152
1.1	230	2.92	6.73	1.429

(a) Board of Gas & Electric Light Commissioners of the Commonwealth of Massachusetts, 1911, pp. 96, 241[a], 222[a]. Acetylene gas companies are not included.

(b) Prices are for 1911 and are computed by using data on p. 226[a] in conjunction with p. 229[a].

Digitized by Google

(B) SUMMARY FOR ALL DATA BY GROUPS OF LOCALITIES.

Size Class of Locality (thousands).	Number of Localities (a) in Class.	Aggregate Population April 15, 1910.	Average Population for Each Class.	Average Price per M cu. ft.	Number of Meters June 30, 1910.	Consumption per Year (Average of 1910 and 1911) (M cu. ft.).	Consumption per Meter (M cu. ft.).	Consumption per Capita (M cu. ft.).
....	1	606,369	606,369	$ .80	128,170	316,518	33.68	7.119
....	1	165,924	165,924	.854	26,560	688,400	25.92	4.149
....	1	145,986	145,986	.806	21,626	675,045	31.21	4.624
120–100	5	567,772	113,554	.862	105,064	2,673,587	25.45	4.709
100– 80	5	449,974	89,995	.907	68,040	1,595,679	23.45	3.546
60– 50	2	116,296	58,148	1.11	18,727	353,292	18.87	3.038
45– 30	5	187,765	37,553	1.06	26,169	580,021	22.16	3.089
30– 20	9	229,294	25,477	1.14	24,538	431,016	17.58	1.880
20– 15	7	118,945	16,992	1.29	12,234	235,425	19.24	1.979
15– 10	9	116,603	12,956	1.47	9,297	160,690	17.28	1.378
10– 6	7	59,861	8,552	1.54	5,043	75,837	15.04	1.267
6– 2	6	28,125	4,688	1.57	1,354	18,423	13.60	0.655
Under 2(b)	1	1,084	1,084	2.92	230	1,548	6.73	1.429

(a) "Locality" means territory served by a particular company so far as population returns permit of such a division or combination of data. The population figures are from the Federal Census as used by the Massachusetts Board.

(b) Not used in plotting the curves.

Digitized by Google

INTERPOLATION OF ORDINATES AND AREAS AMONG AREAS.

By C. H. Forsyth, *Dartmouth College.*

Ordinary interpolation consists geometrically of adding one or more ordinates in a systematic manner to a given series of ordinates. The intervals which separate the given ordinates may be equal or unequal.

Ideal interpolation requires that a large number of the given ordinates shall be used to determine the value of the ordinate or ordinates to be interpolated in order that whatever sense of uniformity exists between the values of the given ordinates shall be maintained after the interpolation. The degree of reliability of the given ordinates will usually decide roughly how many of them should be used to give appropriate results.

The purpose of this paper is to derive a formula to be used to interpolate *areas* as well as ordinates among *areas* in as direct and systematic a manner as ordinary interpolation is and has been performed; however, the areas or ordinates to be interpolated are parts of the given areas and the given areas must be contiguous and of equal breadth.

The use of such a formula can be best explained by a concrete illustration, and the best illustration is to be found in connection with mortality or population statistics, especially the latter because of the great difficulty encountered in attempting to gather such statistics for individual ages. The tendency of human beings to state their ages in figures ending with a "5" or more often a "0" leads to such concentration of data at such ages as to render the results worthless in the form collected. For this reason such statistics are now usually collected in age groups of 5, 10, or some other number of years and then whenever the values for the individual ages are desired they are obtained by interpolation. The method of interpolation used in the past has been very indirect, involving considerable transformation of the data previously. Other

Digitized by Google

problems will readily suggest themselves when once the purpose of the formula is clear.

To illustrate then, suppose the population of a community to be given by the following hypothetical age groups:

Ages.	Population.
10–14	362
15–19	586
20–24	938

If we should then draw a curve in such a way that ordinates dropped at equal intervals would enclose areas which would correspond to the populations given above and in the same order these areas might be thought of as being made up of similar areas which would correspond to the populations for the individual ages. Thus the formula or formulas to be derived may be used to determine either the populations corresponding to individual ages (ordinates) or for any group of ages (areas) which is a fractional part of one of the original groups.

The derivation and application of the formulas to be derived are based upon the calculus of finite differences; however, the theory involved in the application is so simple that anyone can apply the formulas by simply following the examples given in the paper.

If we define the two functions $u_{\underline{x}}$ and $y_{\underline{x}}$ by the relation

$$\Delta y_{\underline{x}} = u_{\underline{x}} = y_{\underline{x+1}} - y_{\underline{x}}$$

whence $$y_{\underline{x}} = \sum_{x=0}^{x-1} u_{\underline{x}} + y_{\underline{0}} = u_{\underline{0}} + u_{\underline{1}} + \ldots + u_{\underline{x-1}} + y_{\underline{0}} \qquad (1)$$

and if we expand $y_{\underline{x+1}}$ and $y_{\underline{x}}$ by the so called Newton's formula in terms of $y_{\underline{0}} (= y_0)$ and its differences, where

$$y_{\underline{x}} = y_0 + \frac{x}{t}\Delta y_0 + \frac{x(x-t)}{t^2}\frac{\Delta^2 y_0}{2!} + \frac{x(x-t)(x-2t)}{t^3}\frac{\Delta^3 y_0}{3!} + \ldots$$

and

$$y_{\underline{x+1}} = y_0 + \frac{x+1}{t}\Delta y_0 + \frac{(x+1)(x+1-t)}{t^2}\frac{\Delta^2 y_0}{2!} + \frac{(x+1)(x+1-t)(x+1-2t)}{t^3}\frac{\Delta^3 y_0}{3!} + \ldots$$

Digitized by Google

we may write

$$u_{\underline{x}}{}_{\underline{i}} = \frac{\Delta y_0}{t} + \Delta \overset{(2)}{F}(x,t)\frac{\Delta^2 y_0}{t^2 2!} + \Delta \overset{(3)}{F}(x,t)\frac{\Delta^3 y_0}{t^3 3!} + \quad . \quad . \quad . \qquad (2)$$

where $\overset{(2)}{F}(x,t)$ refers to the factorial so often found in connection with finite differences and equals $x(x-t)$

$$\overset{(3)}{F}(x,t) = x(x-t)(x-2t)$$

etc.

If we let

$$A = x$$
$$B = x-t$$
$$C = x-2t$$

etc.,

then

$$\Delta \overset{(2)}{F}(x,t) = A+B+1$$
$$\Delta \overset{(3)}{F}(x,t) = AB+AC+BC+A+B+C+1$$

etc.

Formula (2) is the fundamental formula desired but the various terms on the right hand side may be shown to have significant interpretations and meanings.

If we define $w_{\underline{x}}{}_{\underline{i}} = u_{\underline{x}}{}_{\underline{i}} + u_{\underline{x+1}}{}_{\underline{i}} + \quad . \quad . \quad . \quad + u_{\underline{x+i-1}}{}_{\underline{i}}$

then, by (1),

$$y_0 = y_0$$
$$u_{\underline{0}}{}_{\underline{i}} + u_{\underline{1}}{}_{\underline{i}} \quad . \quad . \quad . \quad + u_{\underline{i-1}}{}_{\underline{i}} + y_0 = y_{\underline{i}}{}_{\underline{i}} = y_1 = w_0 + y_0$$
$$u_{\underline{0}}{}_{\underline{i}} + u_{\underline{1}}{}_{\underline{i}} \quad . \quad . \quad . \quad + u_{\underline{2i-1}}{}_{\underline{i}} + y_0 = y_{\underline{2i}}{}_{\underline{i}} = y_2 = w_0 + w_1 + y_0$$
$$y_3 = w_0 + w_1 + w_2 + y_0$$

etc.

Hence

$$\Delta y_0 = y_1 - y_0 = w_0$$
$$\Delta y_1 = y_2 - y_1 = w_1$$

and

$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0 = \Delta w_0$$

Likewise

$$\Delta^3 y_0 = \Delta^2 w_0$$
$$\Delta^4 y_0 = \Delta^3 w_0$$

etc.,

and (2) may be written

$$u_{\underline{x}}{}_{\underline{i}} = \frac{w_0}{t} + \Delta \overset{(2)}{F}(x,t)\frac{\Delta w_0}{t^2 2!} + \Delta \overset{(3)}{F}(x,t)\frac{\Delta^2 w_0}{t^3 3!} + \quad . \quad . \quad . \qquad (3)$$

where it is easy to show that

Digitized by Google

$$\overset{(2)}{\Delta F}(x,t) = 2x - t + 1$$

$$\overset{(3)}{\Delta F}(x,t) = 3x^2 + 3x(1-2t) + (1-3t+2t^2)$$

$$\overset{(4)}{\Delta F}(x,t) = 4x^3 + 6x^2(1-3t) + 2x(2-9t+11t^2) + (1-6t+11t^2-6t^3)$$

etc.

Now, the values w_0, w_1, w_2 . . . are obviously groups of t values, just as we find in mortality and population statistics, and $u_{\underline{x}}$ represents one of these values included in a group, its identification depending upon the value of x. As an illustration let us take the illustration given at the beginning of this paper. Here $t=5$, and $u_{\underline{0}}$, $u_{\underline{1}}$, etc., represent the populations at ages 10, 11, etc., and $u_{\underline{0}} + u_{\underline{1}} + \ldots + u_{\underline{4}} = w_0 = 362$, etc.

Thus

$$\begin{array}{lll} w_0 = 362 & & \\ & \Delta w_0 = 224 & \\ w_1 = 586 & & \Delta^2 w_0 = 128 \\ & \Delta w_1 = 352 & \\ w_2 = 938 & & \end{array}$$

and the population at age, say 15, according to (3) is

$$\begin{aligned} u_{\underline{5}} &= \frac{w_0}{5} + 3\frac{\Delta w_0}{25} - 4\frac{\Delta^2 w_0}{125} + \quad \text{to second differences} \\ &= \frac{362}{5} + 3\frac{224}{25} - 4\frac{128}{125} + \\ &= 95.184 \text{ or } 95 \end{aligned}$$

Since population and deaths are usually given in quinquennial age groups t usually has the value "5." In this case (3) may be written

$$u_{\underline{x}} = .2w_0 + .04(x-2)\Delta w_0 + .004(x^2-9x+12)\Delta^2 w_0 + .0008(x^3-21x^2+116x-126)\frac{\Delta^3 w_0}{3}$$

$$+ .00004(x^4-38x^3+467x^2-2014x+1915.2)\frac{\Delta^4 w_0}{3} + \ldots \quad (4)$$

although we believe it is rarely necessary to use differences beyond the second in such a connection. Of course, in case

Digitized by Google

the formula is needed to only a lower number of differences, the higher differences of the formula are simply ignored.

If x be given the value "7," formula (4) up to and including second differences becomes identical with a formula used by Mr. George King in his method of constructing abridged mortality tables as explained in the report of the English Registrar General for 1914.

So far we have pointed out how formula (3) may be used to determine for example the population for a single age. As a rule, in such a connection, the populations for each and all of the individual ages are desired and in such a case the work may be simplified by using not formula (3) or (4) but the leading term and differences obtained from it when x is given successively the values $\frac{5}{5}, \frac{6}{5} \ldots \frac{10}{5}$ and the results differenced five times. The leading term and differences given in workable form are as follows:

Leading term	$.2w_0$	$+.12\Delta w_0$	$-.032\Delta^2 w_0$	$+.0144\Delta^3 w_0$
1st difference	$.04$ "	$+.008$ "	$-.0064$ "	
2d difference	$.008$ "	$-.0016$ "		
3d difference	$.0016$ "			(A)
4th and higher differences.	0			

As an illustration let us take our original but now somewhat extended problem and break up the population for the age group (15–19) into populations for each and all of the individuals included in the group.

Ages.	Populations.			
10–14	$w_0 = 362$			
		$\Delta w_0 = 224$		
15–19	586		$\Delta^2 w_0 = 128$	
		352		$\Delta^3 w_0 = 105$
20–24	938		233	
		585		
25–29	1523			

When the leading term and differences (the underscored values) are computed, written as shown, and added successively in the ordinary way, we have the following results (shown with the other values occurring in connection with the addition):

Digitized by Google

Ages.	Populations.				
15.....	96.696				97
		9.312			
16.....	106.008		.856		106
		10.168		.168	
17.....	116.176		1.024	0	or 116
		11.192		.168	
18.....	127.368		1.192		127
		12.384			
19.....	139.752				140
	586.000				586

In this illustration the work was carried out to an unnecessary number of decimal places in order to show a perfect check on the work. Usually one or two decimals will prove sufficient, depending upon the size of the values involved.

As a general rule mortality and population statistics, even at their best, are faulty and hence in most cases all third differences may be ignored in which case the leading term and differences may be taken as

Leading term	$.2w_0+.12\Delta w_0-.032\Delta^2 w_0$	
1st difference	$.04$ " $+.008$ "	
2d difference	$.008$ "	(A')
3d and higher differences	0	

It will be noticed that the formulas of leading term and differences given above are applied centrally, *i. e.*, at least one value must be known on each side of the value to be broken up, and hence the formulas can be used in all places except to break up "end" values.

The leading term and differences for breaking up "end" values are obtained by setting $x=\frac{0}{5}, \frac{1}{5}, \ldots \frac{5}{5}$ successively in formula (4) and differencing the results five times and are as follows:

Leading term	$.2w_0-.08\Delta w_0+.048\Delta^2 w_0-.0336\Delta^3 w_0$	
1st difference	$.04$ " $-.032$ " $+.0256$ "	
2d difference	$.008$ " $-.0096$ "	
3d difference	$.0016$ "	(B)
4th and higher differences	0	

Digitized by Google

Thus, the leading term and differences (B) may be used to break up "end" values such as population "362" for the age group (10–14) given above, as follows. However, in this case, only second differences will be used in order to show clearly how the work is finally performed.

Ages.	Populations.			
10.....	60.624			
		4.864		
11.....	65.488		1.024	
		5.888		0
12.....	71.376		1.024	0
		6.912		0
13.....	78.288		1.024	
		7.936		
14.....	86.224			
	362.000			

Here, again, an unnecessary number of decimal places are carried in order to show a perfect check on the work.

Of course, in breaking up the other of the two "end" values occurring in any set of statistics the column of values needs only to be reversed and the formulas applied in the same way. In all cases any negative signs that appear must be carefully preserved and carried along with the work.

The formulas derived above will prove sufficient to cover practically all cases met with in ordinary work of this kind; but, looked at from a geometrical standpoint, the formulas derived so far can be used to interpolate only what might be regarded as *ordinates* among *areas*. To make our paper anywise complete we should include a formula which would allow us in case of necessity to determine, say the population which would correspond to some fraction of a particular age group such as the first 3/5, or the middle 2/10 and, in fact, any fractional part of any group. Such a formula would then be used to interpolate what might be regarded as *areas*.

To establish a formula for determining the value which would correspond say to any $\frac{n}{t}$ part of a given group; it is

Digitized by Google

obviously necessary to merely determine the value of the sum.

$$U_{\frac{nx}{t}} = u_{\frac{x}{t}} + u_{\frac{x+1}{t}} + \ldots + u_{\frac{x+n-1}{t}}$$

in terms of the formula (3), where, to recapitulate, $U_{\frac{nx}{t}}$ gives the value which corresponds to any $\frac{n}{t}$ part of the group with which $\frac{x}{t}$ is connected, where

$\frac{0}{t}, \frac{1}{t} \ldots \frac{t-1}{t}$ fall in the first group,

$\frac{t}{t}, \frac{t+1}{t} \ldots \frac{2t-1}{t}$ fall in the second group, etc.

The summation suggested above is so easily performed by finite integration that merely the final result (to second differences) is given, as follows:

$$U_{\frac{nx}{t}} = \frac{nw_0}{t} + A\frac{\Delta w_0}{t^2 2!} + B\frac{\Delta^2 w_0}{t^3 3!} + \ldots \qquad (5)$$

where $A = n(2x+n-t)$

$B = n\{3x^2+3x(n-2t)+(n-t)(n-2t)\}$

etc.

Suppose, as an illustration, we desire to determine the population of the first 3/5 part of the quinquennial group (15–19) of the example given in this paper, using second differences. Here $t=x=5$ and $n=3$, or

$$U_{\frac{nx}{t}} = 3/5w_0 + 12\frac{\Delta w_0}{25} - 8\frac{\Delta^2 w_0}{125}$$

$$= 316.528 \text{ or } 317$$

which will be found to be the same as the sum of 95.184, 105.168 and 116.176 found separately using second differences.

Digitized by Google

CHARTERS FOR AMERICAN BUSINESS CORPORATIONS IN THE EIGHTEENTH CENTURY.

By Joseph S. Davis, *Harvard University.*

The corporation was naturalized in America in the seventeenth and eighteenth centuries. In the last half century of the colonial era it was utilized, in scattering instances, for business purposes. Immediately after the Revolution a marked extension of this use was one of the numerous politico-economic experiments essayed. The experiment succeeded. Thereupon a business corporation movement, unprecedented in any country, developed in the last decade of the century. The accompanying statistics, based upon a careful if not exhaustive examination of the compiled and session laws of the various states, are presented to reveal the salient features of this movement.*

In the period concerned no classification of business corporations had developed. The charters are here classified, therefore, on a common sense basis, upon which a few comments are in point. The banks were entirely joint stock commercial institutions. Of the 33 insurance companies, 11 were mutual, for protection against loss by fire; of the 22 stock companies, a few were specifically chartered for fire or marine insurance, but the large majority were privileged to write both kinds of risks, while several charters were broad enough to include life insurance, of which practically none was actually written.† The term "highway companies" is used in preference to "transportation companies," since every company here included engaged merely in constructing and maintaining the basis of communication, not in conducting transportation. No distinction between canal companies and others is made in the group of charters for the improvement of inland navigation; for the precise means used varied with the locality rather than with the companies or charters.

*Judge S. E. Baldwin's essay on "Private Corporations" in *Two Centuries Growth of American Law* (*Yale Bicentennial Series*, New York, 1901) has been the authority on this subject. The present statistics are more complete and somewhat more accurate. Additional data, especially relating to individual companies, will appear in the writer's forthcoming *Essays in the Earlier History of American Corporations.*

† Certain ecclesiastical corporations which provided annuities for superannuated clergy or the widows and orphans of clergymen are not included, since they were not business corporations.

Digitized by Google

A few charters gave powers to accomplish varied purposes. Thus a turnpike company was entitled to erect a toll bridge; a manufacturing company had the right to dig canals to afford an outlet for its products; a canal company was empowered to build docks and furnish a city water supply; and a water company was authorized to employ its surplus capital as it should see fit. As a rule, however, one of the objects indicated was clearly the principal one, and such companies have been classified according to the dominant purpose. The White River Falls Bridge Company of New Hampshire and Vermont, however, was apparently equally interested in spanning the river and locking its falls: it has, therefore, been listed under two heads, and to this extent there is a double counting.

Furthermore, several corporations received two or more charters. The most notorious example is the Bank of North America, whose congressional charter of 1781 was buttressed by state charters from Massachusetts, New York, and Pennsylvania in 1782 and Delaware in 1786, and effectually supplanted by a new one from Pennsylvania in 1787. Four companies for building bridges over boundary rivers and 8 companies for improving navigation affecting two states secured charters in two jurisdictions. After making allowance for double counting from these various sources, it appears that not more than 317 distinct corporations were authorized by the 335 charters counted.

A few charters may have escaped the writer's eye. In particular it is probable that several water companies secured charters under the Massachusetts general incorporation act of February 21, 1799; but I have been able to find evidence of none. With this exception, the statistics are believed to be reasonably complete.

The tables here given summarize the principal available information. Table I, in its three parts, gives a summary view of the charters granted, grouped by sections, periods, and general purposes. Tables II and III show the progress of the movement year by year, with reference to sources of corporate powers and the specific purposes authorized, and Table IV roughly correlates these last two features without reference to dates of charters. Table V, finally, brings out certain facts

Digitized by Google

bearing upon the relation of the number of charters to the population in the several states.

Certain characteristics of the movement are evident from a glance at Table I. But 2 per cent. of the eighteenth century charters were granted before the Revolution; 88 per cent. were granted after 1790, and three fifths of these in the last 5 years. The 33 charters granted in 1781–90, moreover, created but 25 distinct corporations. The progress of the movement is further brought out by annual averages for five year periods, as follows:

Period.	Annual Averages.	
	No. of Charters.	No. of Corporations.
1781–1785	2.2	1.2
1786–1790	4.4	3.8
1791–1795	22.6	21.4
1796–1800	35.8	35.2

New England strikingly leads in the number of charters, with 60 per cent. of the total; the others are about equally divided between the Middle and Southern states, if Maryland is counted with the latter. In the decade 1781–90, on the other hand, more charters were granted south of Mason and Dixon's line than north of it; the South shared but slightly in the boom in charter granting after 1790 which so markedly affected New England. The dominant type is clearly the highway companies, which constituted nearly two thirds of the total number; the financial corporations made up 20 per cent., the local public service companies 10, while the business corporations proper added less than 4 per cent. The local public service companies were largely confined to New England, business corporations proper to the New England and Middle states.

The more detailed tables reveal further facts. The years 1791–92 show a rapid increase in new charters. In the single year 1792, more corporations were created than in the entire

Digitized by Google

decade 1781–90. Except Kentucky and Georgia, all of the states had swung into line by 1792, and only Delaware and South Carolina failed to grant at least one charter in that year. The high point of the period is 1795, due chiefly to New Hampshire's large contribution; from this year on, however, a high level is maintained—from the average of 37 per annum there is no deviation greater than 6. Before 1791 nearly half of the charters, and more than half the corporations chartered, were for the improvement of inland navigation; thereafter there was a fairly wide distribution among different types. Disregarding sporadic or experimental charters, one may date the beginning of the *movements* for various types as follows: 1783, inland navigation; 1790, banking; 1791, toll bridge; 1794, insurance; and 1795, water supply.

In order of ubiquity the inland navigation companies come easily first: of this type every New England, Middle, and Southern state chartered at least one. Toll bridge companies appear in 12 states; banking, insurance, and water companies in 10. The leading states in chartering particular types were: *banking*—Massachusetts, with 7 out of 34; *insurance*—Maryland and Rhode Island, each with 6 out of 33; *inland navigation*—Virginia, with 14 out of 74; *toll bridge*—New Hampshire, with 19 out of 73; *turnpike*—Connecticut, with 23 out of 72; *water supply*—Massachusetts, with 15 out of 32; and *manufacturing*—Massachusetts, with 4 out of 8. North Carolina and Georgia chartered none but navigation companies, Kentucky only a bridge company. The predominating type in Maine and New Hampshire was the toll bridge, in Connecticut and New York the turnpike, in Virginia and South Carolina as well as their sister Southern states (except Maryland), the inland navigation company.

Banking and insurance charters were about equally numerous, though insurance companies slightly outnumbered the banks. Highway companies were about equally divided among the three types,—inland navigation, toll bridge, and turnpike. Water companies constituted nearly the whole of the local public service companies, and were about as numerous as banks or insurance companies. Manufacturing companies, few though they were, constituted 60 per cent. of the business corporations proper.

Digitized by Google

The leadership of the New England states is emphasized by the detailed tables. Each ranked higher in number of charters than in population in 1800. The story is the same if one compares percentage of charters with percentage of population. Rhode Island, with 1.3 per cent. of the population, granted nearly as many charters as Virginia, with 16.6 per cent. of the population; Connecticut, with 4.7 per cent. of the population, granted more charters than the four states south of Maryland, which contained 35.2 per cent. of the population; Rhode Island and Connecticut together, with 6 per cent. of the population, granted about as many charters as all the Middle or Southern states. For New England as a whole, 16.3 charters were granted per 100,000 population (1800), as compared with figures for the Middle states of 4.6 and the Southern states of 2.9. In this comparison Rhode Island leads with 28.9 charters per 100,000 population, but backwoods Vermont, the lowest New England state, shows a figure of 12.3 to contrast with 6.1 for Maryland. In each section there is an approach to uniformity in density, although in this respect Maryland fits in best with the Middle states.

The New England states also lead in types of charters. All but one of the various types are represented there, and two of its types (mining, land) are not elsewhere represented. In every type but inland navigation companies, in which the South is ahead, it leads. On the average, each New England state chartered 7 types of corporations, each Middle state, 5, each Southern state, 3. Here Connecticut leads, with 11 different types represented.

Mere statistics of charters are, of course, inadequate. We should like further to know how many corporations went actually into operation, what was the mortality experience, how much capital was employed in all and in each group, what were the characteristic sizes, whether they increased or decreased as the movement progressed, and so on. Records still extant throw gleams of light on these subjects, but nothing approaching complete statistics of them can be hoped for. Yet the few figures here shown indicate a breadth and significance of this early corporation movement which has hardly been sufficiently appreciated.

Digitized by Google

TABLE I.

SUMMARY OF EIGHTEENTH CENTURY CHARTERS TO BUSINESS CORPORATIONS IN THE UNITED STATES, GROUPED BY PERIODS, SOURCES OF CHARTERS, AND GENERAL TYPES.

Sources of Charters	Colonial	1781–85	1786–90	1791–95	1796–1800	Total Charters	*Per Cent.*
United States		1		1		2	*.6*
New England	6	4	4	70	116 (a)	200 (a)	*59.7*
Middle States	1	2	4	22	38	67	*20.0*
Southern States		4	14	22	25	65	*19.4*
Western States					1	1	*.3*
Total Charters	7	11	22	115	180 (a)	335	*100.0*
Per Cent	*2.1*	*3.3*	*6.6*	*34.3*	*53.7*	*100.0*	

General Type	Colonial	1781–85	1786–90	1791–95	1796–1800	Total Charters	*Per Cent.*
Financial	1	5	5	29	27	67	*20.0*
Highway		5	14	79	121	219	*65.4*
Local public service	5			4	27(a)	36(a)	*10.7*
Business (proper)	1	1	3	3	5	13	*3.9*
Total Charters	7	11	22	115	180(a)	335	*100.0*

Sources of Charters	Financial	Highway	Local Public Service	Business (proper)	Total Charters	*Ancillary, additional or Joint Charters*	Total Corporations
United States	2				2		2
New England	33	130	30(a)	8	200(a)	*8*	192
Middle States	16	42	4	5	67	*7*	60
Southern States	16	47	2		65	*3*	62
Western States		1			1		1
Total Charters	67	219	36(a)	13	335	*18*	317
Ancillary, additional, or Joint Charters	*5*	*13*			*18*		
Total corporations	62	206	36(a)	13	317		

(a) Charters to water supply companies issued under the Massachusetts general incorporation act of Feb. 21, 1799, (*Laws*, ed. 1801, ii, 843–847) cannot be found and are not included.

Digitized by Google

TABLE II.

EIGHTEENTH CENTURY CHARTERS TO BUSINESS CORPORATIONS IN THE UNITED STATES, CLASSIFIED BY YEARS AND SOURCES OF CHARTERS.

Sources of Charters	Colonial	1781	1782	1783	1784	1785	1786	1787	1788	1789	1790	1791	1792	1793	1794	1795	1796	1797	1798	1799	1800	Total	Per Cent.
United States		1										1										2	.6
Maine(a)												1	1	2	2	3	3	7	2	1	1	23	6.9
New Hampshire													6	1	2	10	2	4	1	3	3	32	9.5
Vermont												1	1		1	3	1	4		4	5	20	6.0
Massachusetts	1		1		1	1		2		1			8	3	3	5	8	8	6	6	6	60	17.9
Rhode Island	3											1	2		3	1	1			2	7	20	6.0
Connecticut	2				1					1			3			7	3	9	9	2	8	45	13.4
New York			1								1	1	3	1				3	6	7	5	28	8.4
New Jersey												1	1			2	1	2	1	3	2	13	3.9
Pennsylvania	1		1				1	1				1	2	5	3	1	2	1	4			23	6.9
Delaware							1							1			1					3	.9
Maryland				1	1			1			1	2	1	1	1	4	3		4	1		21	6.3
Virginia						2		1	3	1			2	1	1	6	1		2		2	22	6.6
North Carolina									1		2		1			1	5		1			11	3.3
South Carolina							1	2	1						1			3		2		10	3.0
Georgia																				1		1	.3
Kentucky																				1		1	.3
Total	7	1	3	1	3	3	3	7	5	3	4	9	31	15	17	43	31	41	36	33	39	335	100.2
Per Cent	2.1	.3	.9	.3	.9	.9	.6	2.1	1.5	.9	1.2	2.7	9.3	4.5	5.1	12.8	9.3	12.2	10.7	9.8	11.6	100.0	

(a) During this period Maine was, of course, a district of Massachusetts, and charters for her corporations were granted by the Massachusetts General Court.

Digitized by Google

TABLE III.

EIGHTEENTH CENTURY CHARTERS TO BUSINESS CORPORATIONS IN THE UNITED STATES, CLASSIFIED BY YEARS AND PURPOSES.

Purposes	Colonial	1781	1782	1783	1784	1785	1786	1787	1788	1789	1790	1791	1792	1793	1794	1795	1796	1797	1798	1799	1800	Total	Per Cent.
Banking		1	3		1		1	1			1	3	8	3		5	2			2	3	34	10.1
Insurance	1						1	1				1			4	5		3	6	5	6	33	9.9
Financial	*1*	*1*	*3*		*1*		*2*	*2*			*1*	*4*	*8*	*3*	*4*	*10*	*2*	*3*	*6*	*7*	*9*	*67*	*20.0*
Inland navigation				1	1	2	1	3	5	1	2	3	11	5	2	10	11	5	5	3	3	74	22.1
Toll bridge						1		2				1	10	6	7	14	6	14	7	3	2	73	21.8
Turnpike													1		3	6	6	10	11	15	20	72	21.5
Highway				*1*	*1*	*3*	*1*	*5*	*5*	*1*	*2*	*4*	*22*	*11*	*13*	*30*	*23*	*29*	*23*	*21*	*25*	*219*	*65.4*
Water supply	3												1			2	4	8	6	4	4	32	9.6
Dock	2															1			1			4	1.2
Local Public Service	*5*												*1*			*3*	*4*	*8*	*7*	*4*	*4*	*36*	*10.7*
Manufacturing										2	1	1			1		1	1			1	8	.3
Mining					1																	1	.6
Agriculture														1								1	2.4
Land																	1					1	.3
Commercial	1																			1		2	.3
Business (proper)	*1*				*1*					*2*	*1*	*1*		*1*	*1*		*2*	*1*		*1*	*1*	*13*	*3.9*
Total	7	1	3	1	3	3	3	7	5	3	4	9	31	15	17	43	31	41	36	33	39	335	100.0
	2.1	*.3*	*.9*	*.3*	*.9*	*.9*	*.9*	*2.1*	*1.5*	*.9*	*1.2*	*2.7*	*9.3*	*4.5*	*5.1*	*12.8*	*9.3*	*12.2*	*10.7*	*9.8*	*11.6*	*100.0*	

Digitized by Google

TABLE IV.

EIGHTEENTH CENTURY CHARTERS TO BUSINESS CORPORATIONS IN THE UNITED STATES, CLASSIFIED BY PURPOSES AND SOURCES OF CHARTERS

Sources of Charters	Financial		Highway			Local Public Service		Business (proper)		Total
	Banking	Insurance	Inland Navigation	Toll Bridge	Turnpike	Water Supply	Dock	Manufacturing	Miscellaneous	
United States	2									2
Maine	1	1	7	12		1	1			23
New Hampshire	1	1	5	19	4	1	1			32
Vermont			5	5	9	1				20
Massachusetts	7	5	5	14	9	15	1	4		60
Rhode Island	4	6	1	3	3	3				20
Connecticut	5	2	2	3	23	5	1	1	3	45
New York	4	3	3	1	13	2		2		28
New Jersey			4	5		2		1	1	13
Pennsylvania	3	4	5	5	5				1	23
Delaware	2		1							3
Maryland	3	6	4	4	3	1				21
Virginia	2	3	14		3					22
North Carolina			11							11
South Carolina		2	6	1		1				10
Georgia			1							1
Kentucky				1						1
Total Charters	34	33	74	73	72	32	4	8	5	335
Ancillary, additional, or Joint Charters	*5*		*8*	*4*						(a)*18*
Total Corporations	29	33	66	69	72	32	4	8	5	317

(a) One bridge and canal company occasions an additional subtraction.

Digitized by Google

TABLE V.

COMPARISON OF STATES WITH RESPECT TO POPULATION, 1800, AND NUMBER OF EIGHTEENTH CENTURY CHARTERS TO BUSINESS CORPORATIONS.

Sources of Charters	No. of Charters Granted	No. of Charters per 100,000 Population 1800	Percentage of Total Charters Granted	Percentage of Total Population 1800	Rank with Respect to No. of Charters Granted	Rank with Respect to Population 1800
United States	2		.6			
Maine	23	15.2	6.9	2.8	5–6	14
New Hampshire	32	17.4	9.5	3.5	3	11
Vermont	20	12.3	6.0	2.9	9–10	13
Massachusetts	60	14.2	17.9	7.9	1	5
Rhode Island	20	28.9	6.0	1.3	9–10	16
Connecticut	45	18.0	13.4	4.7	2	8
New England	*200*	*16.3*	*59.7*	*23.2*	*1*	*3*
New York	28	4.8	8.4	11.1	4	3
New Jersey	13	6.2	3.9	4.0	11	10
Pennsylvania	23	3.8	6.9	11.3	5–6	2
Delaware	3	4.6	.9	1.2	14	17
Middle States	*67*	*4.6*	*20.0*	*27.6*	*2*	*2*
Maryland	21	6.1	6.3	6.4	8	7
Virginia	22	2.4	6.6	16.6	7	1
North Carolina	11	2.3	3.3	9.0	12	4
South Carolina	10	2.9	3.0	6.5	13	6
Georgia	1	.6	.3	3.1	15–16	12
Other Southern States				.5		
Southern States	*65*	*2.9*	*19.4*	*42.0*	*3*	*1*
Kentucky	1	.4	.3	4.2	15–16	14
Other Western States				3.0		
Western States	*1*	*.3*	*.3*	*7.2*	*4*	*4*

Digitized by Google

THE APPLICATION OF STATISTICS TO ADVERTISING AND MARKETING.

By Hyman L. Roth, *Statistical Bureau, Metropolitan Life Insurance Company.*

When one mentions the term statistics, people are wont to associate with it the economist, the social worker, the government official, the college professor, or perhaps the insurance statistician, the financier, and the like. But the business man engaged in the manufacture and sale of some product, is rarely thought of in connection with the uses of statistics. And yet, statistical data and research methods have assumed vital importance and have been extensively used in various phases of business life. Business executives are not only searching for and utilizing the figures compiled by the national and state governments and other bodies, but are also themselves gathering and tabulating statistical data germane to their business. This is particularly true in the fields of advertising and marketing, to which this paper will be devoted.

The uses of various kinds of statistical data for the determination of markets, the application of statistical methods for checking up results, for apportioning sales quota, for measuring competition, for laying out advertising plans and selling campaigns—*specific instances* of how statistical methods and data have been used for these and other purposes will be presented below.

Attitude of Business Men toward Statistics. Nor does the business man refer to statistics in a slighting manner. On the contrary, he is attaching more and more significance to the possibilities and utility of statistics. Let me quote several items to bring out this point.

In an address before the teachers of advertising at the 1916 convention of the Associated Advertising Clubs of the World, S. Roland Hall, advertising manager of the Alpha Portland Cement Co., quoted a number of passages from letters that he had received from advertising and sales executives in answer to the query as to what they expected from students who

Digitized by Google

completed courses in advertising. Among the replies was the following:

They (those who have studied advertising) are inclined to emphasize the copy side and entirely overlook the statistical and research work that is an indispensable preliminary. I think that teachers might do well to teach students the usefulness of graphs and charts, train them to investigate, to ascertain populations in given groups, sales possibilities, etc. They should be shown how salesmen are routed by maps, and how selling effort generally is organized and directed.

The following passage, from an editorial in *Printers' Ink* of March 30, 1916, a weekly magazine for advertisers, is also indicative of the same attitude.

Uncle Sam is an indefatigable gatherer of statistics, and the publications of the Departments of Commerce and Agriculture are veritable mines of facts and figures which have been compiled at the public expense, and which are supposed to be for the public benefit. The usefulness of all this matter, however, is nothing like what it might be if it were properly brought to the attention of those who might make profitable use of it. Unless a business man himself takes the initiative, and writes to Washington for information, he is not likely to appreciate the broad scope of the work which is actually being done for him by the government.

Then, referring to a booklet issued and circulated by the Census Bureau, directing attention to publications of a statistical nature of possible use to business men, the editorial concludes:

We are glad to note that the need (that is, for publicity of statistical data) has at last been recognized by one government official at least, and we hope the practice will spread.

Another interesting indication of the attitude regarding statistics for business purposes, is the growing use of statistical terminology. Thus, in *Associated Advertising*, February, 1916, an article contains the phrase, "a movement which looks toward reducing the business death rate" etc. Again, in the March, 1916, issue of the same magazine, there is an editorial on the efforts of the Ingersoll Cost Accounting Committee of the Associated Advertising Clubs of the World, containing such phrases as the "high business death rate" and "when the

Digitized by Google

desired *figure facts* are obtainable, it will be possible to arrive at various *safe averages* for income and outgo," etc. In an interview reported in *Printers' Ink*, April 15, 1915, W. H. Ingersoll is quoted as saying: "The vital statistics of a business or department, like the vital statistics of a nation or government, may best be concentrated for preservation in the form of graphic charts." While passages of this kind can be said to be incidental, yet they are significant of the attitude of business men on this subject.

The general manager of a concern manufacturing talcum makes the following comment on the lack of statistics in that field as compared with other lines of business, *Printers' Ink*, Feb. 3, 1916:

But in the talcum business, it has been altogether different. There have been no statistics, no appreciation of the real and possible demand, no accurate knowledge of per capita consumption. We are all of us, even the largest, losing far more by secrecy than we would gain by a free interchange of data and opinion.

At a meeting of the Advertising Club of New York, in March, 1916, John Lee Mahin proposed what is really the statistical method for measuring the "good will" produced by advertising. He said, as reported in *Printers' Ink*, April 6, 1916:

First we can measure the amount of money spent over a period, say, of three years. We can take the amount of space and picture it by the graphic method. Then we can put against it the total amount of space it bought. We can cut that space up and say that a certain large percentage of it has been put into the propagation of the ideas. And that another smaller portion of the space, say, 5 per cent., has been devoted to making the street location of the house known; and 2½ per cent. to a name that fits the tongue easily or a mark to identify it; and 5 or 10 per cent. more to more or less inconsequential things.

I say we could do that with more or less facility. We could do it also with competitive articles and check the good will of these against the good will of the goods in question.

But that is not all. We could then send out into the field and canvass large enough groups of typical consumers and those who have influence on consumers and in this way get an *accurate approval of the state of mind* of those consumers and confidently believe that they are representative of the whole

Digitized by Google

mass of consumers. We should never lose sight of the fact that the opinions, desires, preferences of the consumers are *facts.*

Just one more reference of this kind, an editorial from *Printers' Ink,* June 29, 1916:

Indeed, the whole trend of the advertising business is straight away from the old methods of trial and error, and toward an even greater dependence on facts and figures. Nowhere is that fact more apparent than in the increasing use of statistics, which often indicate the advisability of a change in tactics before it becomes painfully evident in the sales total.

The foregoing passages are quoted to indicate, at the very beginning, the seriousness with which business men regard statistical data and methods as applied to their respective fields. Let us now pass to a consideration of a number of *actual instances* of the uses of statistics by the three following elements of the field of advertising—the advertiser, the advertising medium, the advertising agency. In citing the various instances, I have preferred to have those who have utilized statistical methods tell their own story in their own words, and have therefore quoted rather extensively.

USES OF STATISTICS BY ADVERTISERS.

The use of statistics by a manufacturer is well exemplified in the case of Robt. H. Ingersoll and Bro. in an interview reported in *Printers' Ink,* April 15, 1915:

In these days of aspiration for scientific management every executive is trying to get his business on a measurable, analyzable basis. The original unit of comparison for determining this almost always has to be the productiveness of the individual salesmen. Afterwards the measuring-stick is the territory. The Ingersolls are now coming to the third standard, units of population; a variation of the territorial idea, but yet a variation. The company now sells 46 or 47 watches per thousand of population; two or three per thousand in secluded sections, 100 or more per thousand in cities, but about 46 or 47 per thousand the country over. But how much *ought* it to sell?

To know the number of watches sold per thousand and the cost per thousand gives a much broader and sounder basis than the other way. We ask how many can be sold? The life of the average watch is so many years, and in that we

Digitized by Google

include breakage and loss. Replacement is so much, competition is so much. The relations of wealth and poverty, of factory and commercial and agricultural communities, are all figured in. Averages are found for state, sections and the country as a whole.

With these in hand we can turn around and make local investigations with entirely new and highly important resources of inquiry. Why are the sales per thousand greater in this place than in that? Why should it cost more to make fewer sales there? Territories show up in new light, in better focus, and we can see whether the need is primarily for one kind of help or for another, this kind of appeal or that.

So important do some people consider statistical research in its application to business problems, that they believe in the existence of a separate department for the sole purpose of conducting work of this kind. Thus, Chas. C. Parlin, Manager of Division of Commercial Research of the Curtis Publishing Co., in an article in *Printers' Ink*, October 22, 1914, entitled "Why and How a Manufacturer Should Make Trade Investigations," advocates the creation of research departments by manufacturers. A few paragraphs from this article will indicate the utility of statistical data for the purposes indicated:

Today the supply in most lines has caught up with the demand and a third very important function devolves upon the manufacturer; namely, to develop his markets. This function involves first of all a thorough knowledge of his existing markets and of all those influences which are operating to affect them. It seemed natural enough in the earlier stages of industrial development for the manufacturer to confine his attention to the making of goods and to entrust to an outside sales organization the second function of getting the goods to the consumer. It is still advantageous in many lines for the manufacturer to reach the retailer through jobbing connections, but no manufacturer, however efficient and honorable the middlemen handling his product are, can afford to be without firsthand knowledge of his markets. Every manufacturer should know where his goods are sold, who buys them, and why they are bought, what type of men are selling his goods to consumers, what influences are affecting them, what their sales methods and sales costs are, to what extent they are real factors in making sales and to what extent they are only order-takers.

Digitized by Google

Every manufacturer should know whether he has a uniform distribution or whether there are certain sections and communities within sections where sales are subnormal; many manufacturers reaching the retail trade through middlemen seem to have very little of this information.

Mr. Parlin states further:

A certain manufacturer in the central west was interested primarily in breaking into the New York markets; research showed him that totally neglected at his own door there lay a larger market easier to get and likely to prove more profitable than the coveted New York market; research showed another manufacturer that his distribution was far from uniform, another that he was restricting his line to jobbers when the possible sale for his goods was almost confined to those stores which aimed to buy direct; another who sold only direct that a major portion of the opportunities in his field could be best reached through jobbing channels.

As regards the scope of commercial research, Mr. Parlin states:

Research work should not only be dominated by honesty of purpose, but it should be conducted from the student standpoint of truth for truth's sake. Lines of inquiry that appear likely to prove of practical value should, of course, be pursued. But that which appears academic should not be neglected, for it frequently happens that what appears to be academic turns out to be highly practical, while something which appears likely to be practical turns out to have little value. . . . Commercial research work involves three distinct phases: First, library work; second, field work; third, formulating conclusions.

The first step naturally is the study of printed sources, such as the census reports and other government publications. The research department may often throw light upon a business by merely graphing census figures. Often we have found a manufacturer with the latest census reports at his elbow, but deeply interested in maps and graphs of the same census material as it applied to his own business, indicating, that while he had perhaps read the figures, they came to him in a new light as he saw them graphed.

In advertising and selling, as in vital and social statistics the importance of reducing figures to a rate per hundred or thousand basis has been recognized. Says C. C. Casey in an article, "Graphic Sales Pictures That Analyze the Business," *Printers' Ink*, January 20, 1916:

Digitized by Google

Many of the big sales organizations use the "quota" system in compiling certain of their sales and advertising statistics, and these figures remove some of the inequalities of territorial or agency sales.

The quota in Arizona, may, for instance, be ten sales a month and in New York ten times as many, but the tabulation would show, not the *sales*, but the *per cent. of sales to quota*—75 per cent. of quota in each state would indicate equal comparative results, even though it did only mean that Arizona made 7½ sales and New York 75 sales.

The most graphic method, however, of showing sales or advertising by territories, or by states, or by salesmen is the "per prospect," "per hundred prospect," or "per thousand prospect" method. . . .

If you say that the Arizona sales organization—whether it be salesmen, jobbers or dealers—sold 333 units of your product per thousand prospects in a year and New York sold 111 units, the comparison has a "Jim Jeffries punch" which is likely to jar somebody into demanding a reason.

Another instance of the use of statistics and the rate method was recited by N. C. Kingsbury, Vice-President of the American Telephone and Telegraph Co., at the 1916 convention of the Associated Advertising Clubs of the World, reported in *Printers' Ink*, June 29, 1916. Mr. Kingsbury said:

Our national advertising campaign began in June, 1908. . . . For the five years of 1904 to 1908, inclusive, we gained 1,690,078 subscribers, and this gain was 1.72 telephones for each hundred of the total population of the United States. That was before our advertising campaign began.

Now let us take the period from 1908 to 1913, inclusive, omitting, as you will note, the years 1914 and 1915 in order to avoid the effect of abnormal conditions due to the European war. During these five years, while our advertising campaign was in progress, we gained 2,199,964 stations, and that gain was 1.95 telephones per 100 of the total population of the country, so that during the period covered by our national advertising the gain in telephones was .23 of 1 per cent. greater than during the period when we were not advertising. If you will apply this .23 of 1 per cent. to the total population of this country you will notice it represents a large number of telephones. And this gain was made in spite of the fact that the possibilities for new business were considerably less in the latter period than in the former period.

Another very interesting instance of the use of statistics

Digitized by Google

was described in an article entitled "Testing the Value of Inquiries," also by C. C. Casey, in *Printers' Ink*, January 27, 1916. The advertising of a firm manufacturing machinery was adversely criticized by the management on the score of poor sales results in following up the inquiries with salesmen. So critical did the matter become that the discontinuance of further advertising was being considered. Whereupon the advertising manager made a thorough statistical analysis of the records of inquiries and sales, and presented his results in both tabular and graphic form.

The *length of the sales lines*, showing sales per thousand possible customers, *followed the length of the advertising lines*, showing inquiries per thousand possible customers.

When inquiry lines were long, showing a large number of inquiries per thousand possible customers, sales lines also were long, and *vice versa*—in a manner that was graphic.

There were many variable elements, of course—greater efficiency of men, better crop conditions, and a lot of other things—but the *tendency* was for sales to "grow thick" exactly in proportion as advertising inquiries were "planted thick."

In other words, even though salesmen ignored inquiries, their prospect list as a whole was educating itself to buy, and was buying freely or rottenly in proportion to the number who could be induced to write the company for more information.

Another instance of the use of statistical analysis for determining the value and nature of inquiries, was presented in a talk before the Adcraft Club of Detroit by Edwin A. Walton, Advertising Manager of the Burroughs Adding Machine Co., reported in *Printers' Ink*, Nov. 4, 1915:

Our analysis of our business experience, covering hundreds of different letters, folders, magazine advertisements on every phase of our business, proves definitely and conclusively four things:

First: The more educational our message, the fewer the inquiries.

Second: The more Burroughs machine we talked, the fewer the inquiries.

Third: The more mere curiosity we excited, the greater the number of inquiries, and

Fourth: The more valuable looking the free booklet we offer, the greater the number of inquiries.

Digitized by Google

Statistics as a Guide for Copy. The realization of the above facts gave the Burroughs Adding Machine Co. also a splendid guide as to the kind of copy that was best suited for their purposes. Again, to quote Mr. Walton:

Now we are reasoning this way in our advertising, that if the prospects in the country, numbering two million or more, really knew what a Burroughs would do for them in the way of saving time, insuring accuracy, and giving them without any cost the facts and figures of their business, on which to build better profits—if they really *knew* this, they would be clamoring to buy. They would mob our hundred and seventy offices around the country. They would be sitting on the doorsteps in the morning waiting to buy machines. Now that's the truth. If they only *knew*. We believe that. It is, then, an educational problem.

A further excellent instance of the use of the statistical method for obtaining a guide for copy, was the recent prize contest on the uses of Bon Ami. Says R. S. Childs, general manager of the Bon Ami Co., in an article in *Printers' Ink*, August 10, 1916:

The prize contest had considerable value, however, in fulfilling its *primary purpose of business research.* The traditional theory of the uses of Bon Ami, on which our advertising copy rests, is based on very illusive data. . . .

By tabulating the replies we produced a big chart, showing the seventy leading uses for Bon Ami and to some extent this chart guides our advertising. It leads us, for example, in an advertisement showing the cleaning of a bathtub to put a can of Bon Ami powder, rather than a cake, into the hands of the model, whereas in a window- or mirror-cleaning advertisement we take pains to show a cake, there being a clear preference on the part of housewives along these lines in the prize-contest returns.

For Figuring Advertising Appropriations. The statistical method is used extensively in determining advertising expenditures and appropriations, especially with regard to the localities and mediums in which to advertise. For example, Lee Anderson, in an article, "Some Suggestions on Handling the Advertising Appropriation," *Printers' Ink*, June 22, 1916, recites an instance of the study of a certain advertising problem in which he found that, "with 65 per cent. of his distribu-

Digitized by Google

tion in towns of 5,000 and under, the advertiser was spending 85 per cent. of his appropriation in publications whose circulation was in cities of 25,000 and up."

The following is another instance of the use of the statistical method in the consideration of the advertising appropriation of a national advertiser. To determine how much should be appropriated by this firm for advertising in its various territories, it was proposed to make a test over a period of two or three years by running advertisements in certain territories operated by this firm, and by not advertising in other territories, and then checking up the results. The article describing this analysis, "The 'Saturation Point' in Advertising Expenditure," by C. C. Casey, in *Printers' Ink*, September 9, 1915, is here quoted as a particularly good illustration of the intensive use of statistical methods:

In an analysis, however, to find the right territories, it was discovered that some territories already had been receiving more advertising than others, both in the magazines, because of varying circulation, and in the varying attitude of mind of the different branch managers toward advertising.

When this discovery was made, the territories were sorted into six groups, according to the amount of advertising done in each. The amount of advertising was at first taken as the number of prospective customers reached (and the number of times reached) both by mail and through the magazines.

This, however, was not found satisfactory because of the varying kinds of advertising used and because of variation in circulation of the magazines. . . . . A better method was found in taking the *inquiries* received from the advertising for each territory, ignoring those received direct by the branch managers, that is, those not coming through the home office. . . . .

Because of the varying number of inquiries and the varying number of prospective customers in the different territories, it was decided to reduce the inquiries per territory to the "inquiries per thousand prospective customers" in each territory. This was done by dividing the total number of inquiries in each territory by the number of thousand prospective customers in each territory. . . . .

The territories were then regrouped into six groups, according to the amount of advertising done. In group 1 were put all agencies with less than 100 inquiries per thousand prospective customers; in group 2, all who had between 100 and 150;

Digitized by Google

group 3, all between 150 and 200, and so on until group 6, which had all territories that had more than 300 inquiries per thousand prospective customers.

Then the total sales were taken for all the agencies in each group, and these being divided also by the number of thousand prospective customers in the group of territories, gave the *sales* per thousand prospective customers in each group.

Then the amount of these sales was obtained by multiplying by the average sale price and the total in dollars was used as a basis to estimate percentage of the advertising cost.

Enough of this article has already been quoted to show the intensive use of statistical methods in the case in question. The following paragraph from a series of articles in *Printers' Ink*, in 1915, is also indicative of the application of statistics in determining the advertising appropriation:

It is a fact, however, that the ratio between advertising expenditures and gross sales, when it can be obtained in the form of an average for the entire industry or the major part of an industry, may be very useful. The concern which is able to approach the subject of an advertising appropriation with the knowledge that the normal or average advertising expenditure in its field is 7 per cent. of its gross sales will find its problems materially simplified. Its analysis of the market with respect to its own product will indicate wherein that normal average must be exceeded, and wherein it may reasonably be expected to prove adequate.

George F. Lord, Manager Advertising Division, Du Pont Powder Co., in discussing the problem of advertising appropriations in *Printers' Ink*, March 4, 1915, states:

I am convinced that more failures in advertising are due to lack of analysis of conditions than to any other reason. For lack of such analyses many concerns that sadly need advertising fail to use it, some advertise who should not, and many advertise "too much, too quickly," creating sales demand they cannot meet.

An analysis of a business along the lines here suggested will often develop unthought-of possibilities of expansion or obstacles to success that must be surmounted or circumvented before an advertising campaign can succeed.

He then presents a very comprehensive list of items, the filling in of which would involve the utilization of statistical

Digitized by Google

data already compiled, or the conducting of new investigations to obtain the necessary information.

Studying Competitive Advertising. A most novel and interesting example of statistical analysis is contained in a study of the competitive periodical advertising and literature of several firms in the same field, described in an article in *Printers' Ink*, October 14, 1915, entitled "Sizing Up the Other Fellow," by S. W.:

To get down to practical issues, the analysis divided itself into three parts. One took up the advertising, national and trade, of all the companies in the field; the second regarded the sales-promotion booklets sent out from various offices in response to inquiries of dealers and customers and distributed through dealers, jobbers and salesmen. This included sales promotion methods. The third part concerned the talking points used by various other companies in putting forth the advantages of their respective chests.

The work of finding out about the national advertising of competitors was simple. This part of the analysis came, almost completely covered in detail, from the offices of various magazines willing to give the data. It resolved itself into securing the space requirements of each company as the information was given by the leading national magazines.

The data thus furnished was worked up in chart form, of which the author of the article states:

The value of such a chart, since practical value is the real object, could best be brought about and "cashed in on" after a second year. Comparative statements are always more valuable than simple figures. By comparing what a company has done one year with what it did the preceding year, a person can get a fair idea whether the company is increasing or decreasing appropriations and thereby judge, in a crude but somewhat definite way, whether the company has gone ahead or behind.

Charts were also worked up for the second part of the investigation, which dealt with booklets, correspondence, and other sales-promotion methods of competitive companies:

The use of such a chart is simple: it shows at a glance, how much follow-up work other companies are doing, how much material they send out and the interval of time between letters. Then, it can be judged with some accuracy which of the companies are enjoying success in proportion to the efforts put forth. . . . .

Digitized by Google

Two such charts, one for each of two successive years, would enable a company to judge which of its competitors had particularly increased its efforts in such direction and so a company would be able to tell something of the progress in organization and promotion work made by competitors. So it would happen, as years passed, that these charts would increase in value, each one of them showing up some change in the methods of various companies: then as one or two of the companies might advance far ahead of the others, or drop back, the reasons could be conjectured with a degree of exactness. This, indeed, is a laboratory method of investigation.

As to the third part of the investigation, dealing with the talking points used by the various companies, the following passages will be of interest:

After taking each company separately and listing its talking points and the number of times each point has been used in folders, booklets, national advertising, and miscellaneous respectively, then it was advisable to have a summary made of all the talking points regardless of the companies using them. . . . .

The final part of the analysis made under section three was to give in simple, terse manner the exact methods of approach of various companies. . . . . Methods of attack and appeal were thought out with newer vision after that investigation was completed.

As for the results and value of this investigation, the following conclusion speaks for itself:

It was practical in that it put into usable form what all the competitors in the field were and are doing in way of sales-extension work, sales-promotion work, and efforts to get new and repeat business. The investigation developed in itself a good bundle of notes and ideas and plenty of new suggestions as to methods for reaching and coöperating with dealers, and some eye-opening facts regarding merchandising methods and principles.

Analyzing and Locating Markets. Statistical data and investigations have also proved of much value for purposes of analyzing and cultivating present markets, and for discovering and locating new markets for merchandise. The instances already cited of how the Ingersoll firm uses statistics and the suggestions of Mr. Parlin are good examples. Several other cases of interest may also be given in this connection.

Digitized by Google

An interview with F. D. Gould, Retail Sales Director of the Regal Shoe Co., reported in *Printers' Ink*, April 26, 1916, refers to a certain change of plan in the general sales policy of that company. The following paragraph from that article is significant of the use of statistics in dividing the country into districts or zones.

Immediately it became apparent that any plan of this kind could not be operated from the Boston office. So, after an analysis of conditions, the country was divided into fifteen zones, following the distribution of the per capita wealth of the country, population, and so far as possible, the lines of present Regal district boundaries.

The following quotation from an article in the June, 1916, issue of *Associated Advertising*, entitled "Selling Lessons from the Shop," by Edward S. Babcox, Advertising Manager of the Firestone Tire and Rubber Co., is also illustrative of the use of statistical data for market analysis:

In a certain city the other day, I talked with a man who was erecting a wonderful garage. When I first saw him he was planning to take care of several thousand motor trucks. He was surprised when I showed him after a little analysis and investigation, that there were only 600 or 700 trucks in his whole territory, and that he could not hope to get business from more than 10 to 20 per cent. of these. These facts caused him to alter his plans.

In a big national marketing problem, the sales manager must analyze his field just as carefully. The United States Census figures provide unlimited information and enable him to assign sales quotas to each territory in an intelligent and accurate way.

He knows, for example, that there are about 5,500,000 families in the United States having an income of $1,200 and about 3,250,000 families with an income of $1,800 or more. If we consider $1,800 the minimum income for an automobile-owning family, statistics prove that the field is 59 per cent. developed. If you work on the $1,200 basis, the field is 35 per cent. developed.

The live sales manager knows there are today about 1,750,000 automobiles in the United States. He knows there are 10,000 in St. Louis, for example. He knows there are 23,295 people in St. Louis who can afford an automobile, and he knows that about 43 per cent. of these possible automobile buyers now own one or more cars. With these figures to guide him, he can direct successful selling efforts in St. Louis.

Digitized by Google

The American Radiator Co. is another concern that has utilized statistical data to excellent advantage in planning local advertising campaigns in certain territories. Louis Bruch, Vice-President of the Company, gives an interesting statement in this connection, in *Printers' Ink*, April 8, 1915:

Before we will consider putting on a campaign in a certain city there are many facts we must know. We want to know how prosperous the community is. For example, we are now considering adding papers in cities dependent on potteries. Potteries are very busy these days—busier than ever before—on account of the favorable tariff and war conditions. We want to know what the bank clearings are; if the locality has any advantages in the way of better soil or crops, mining property or other things which make for continued prosperity. We want information as to the make-up of the people so that we may know if they are of a type that will respond to our advertising; what kind of farming is done in the surrounding country; all the figures and facts available about the earning power of the residents; what kind of homes they have, and other data of a kind which will help us to determine in advance whether or not it would pay us to put on a campaign.

Proctor & Gamble also base their sales policies on actual market facts, ascertained by a study of conditions as they are in the field. In an article by S. C. Lambert, in *Printers' Ink*, April 27, 1916, appears the following passage:

There is the widest possible variation in the methods of marketing Ivory Soap and Star Soap, for example, and Crisco requires a still different combination of methods. Each campaign is based upon a study of conditions as they actually are, not as they were forty years or so ago when Ivory Soap was a new brand. This constant and intensive study of actual conditions, with the resulting adaptation of the selling campaign to fit the conditions, are the features of the company's work which have greatest interest for manufacturers in other lines.

A final instance that I shall cite of how the statistical method was used in analyzing a market is the case of Novo Engines, described in an article, "Saving a Business by Switching to New Market and Plans," *Printers' Ink*, May 27, 1916. In this case an analysis brought out the advisability of entering the city market, whereas theretofore only the rural territory was cultivated.

Digitized by Google

It was found when the investigation had been completed that there was a variety of uses to which the new gasoline-engine might be put in the city. There was hoisting in the building and contracting trade. The big hoisting jobs on new buildings were done by big steam-engines or electric current. But there were hundreds and thousands of small hoisting jobs, too small to call for a large steam-engine, too small to warrant the whole time of an engineer and assistant, and too expensive when done by hand labor. . . . .

Then there was the pumping to be done by builders, contractors, and engineers, such as getting the water out of a cellar or a ditch. One contractor paid four men two dollars a day each for several days to empty a ditch of water, and the Novo Engine with pump could empty in less than a day at a cost of thirty cents.

An Intensive Study of Retailers. R. P. Spencer, of the American Optical Co., gives the following account, in an article entitled "A System to Stop Guessing in Dealer-relations," *Printers' Ink*, May 18, 1916, of an exhaustive study that he made along the lines indicated by the title:

Some years ago, when I was merchandising tinware and enameled ware, it occurred to me that our salesmen knew a great deal about our retailers and our prospects that would be very helpful if we had it on file in the office. Such information could be used in many ways—in writing the individual retailer, in getting up circular letters, as guidance in establishing policies and in the successful handling of 101 details. On considering the subject further we saw that, while the salesman already possessed much knowledge about the retailer and his town, he could easily gather a great many other useful facts for us by asking a few questions and by doing a little observing. Charting these data on cards would give us a comprehensive bird's-eye view of the dealer, his store, his clerks and everything else that vitally concerned him. To get the required information we prepared a blank containing sixty-three questions. Our salesmen were requested to fill out one of these blanks for each merchant on whom they called. Of course all these questions were not asked in one trip, but answers were gradually compiled in several visits.

Then follows a list of the sixty-three questions which constituted *in toto* a most searching analysis of the conditions and merchandising methods of the retail trade. As regards the value of this analysis, the following passage is of interest:

Digitized by Google

Supposing we were planning a trade-paper campaign. By getting the answers from our cards to questions numbers 1, 2, 3, 4, and 5 we could know what publications to use and have a fair idea of the kind of copy to insert. If we were looking for new talking-points, the answers to question 10 would give us a mine of material. Answers to those questions about what our competitors were doing furnished us with a lot of valuable tips, both for the improvement of our products and of the methods used in marketing them. How could we prepare consumer advertising intelligently without referring to those cards again and again? In a word, how could we do anything, and be sure that we were right without consulting the facts recorded in our files? After we started this system and began to realize how valuable the information was, we saw that we had always been working in the dark before and had been guessing at things we should have known.

A similar instance of obtaining data regarding retailers is that of the Westinghouse Co., described by R. E. Dildine in *Printers' Ink*, May 25, 1916:

The advertising department has on file a salesman's report on every distributor, which contains detailed information that is very valuable as a basis for planning and distributing dealer-help material.

These data include, for example, such details as the number and size of the dealers' windows, size and location of the store, the class of customers, the number of clerks employed, the mediums of local advertising employed and the kind of dealer-helps that are used and the way they are used.

These reports are a check on waste of advertising materials, as the company furnishes all of the usual selling-helps, such as street-car cards, posters, booklets and newspaper cuts.

Miscellaneous Uses by Advertisers. James Melvin Lee, Director of the Department of Journalism of New York University, sent a letter to a number of firms, and asked among others, the two following questions:

If you had at your disposal an investigator who was well qualified to dig for helpful facts, at what leading tasks would you set him?

What class of information has proved to you to be the most helpful in the past in enabling you to get results from your advertising?

The replies that he received, presented in an address before the American Association of Teachers of Journalism and re-

Digitized by Google

ported in *Printers' Ink*, April 27, 1916, are full of suggestions regarding the actual and possible uses of statistical data and research work.

J. M. Campbell of the Proctor & Gamble Company is quoted as saying:

Not one business house in a thousand knows the facts about its product—knows what there is about it that appeals to the public.

I happened to know of a very large manufacturer of a product which is used in a very considerable percentage of the homes of the United States. He has spent millions of dollars in advertising the fact that this product has such and such a quality. A house-to-house investigation covering 600 homes and an investigation through the mail covering over 1,000 women brought to light the fact that only one person had been influenced by the repetition for the last twenty years of that special feature of this advertiser's advertising. On the other hand, 80 per cent. of the women who answered our inquiry expressed themselves as wanting this particular advertiser to do such and such a thing, something which he was not doing and did not know his public wanted him to do.

R. P. Spencer, Sales Manager of the American Optical Co., stated that he was actually going to give to an investigator the following assignment:

I intend to make arrangements with a trained investigator to secure important facts which will enable us to plan a comprehensive sales-advertising campaign. In the first place, he will arrange with schools, factories, stores—in short, business institutions where work varies—to find out first, how many people are actually wearing glasses at the present time. Second, how many people should have glasses. Third, whether applicants for positions in these places have to pass any eyesight test or answer any questions in regard to it. Fourth, whether an oculist is regularly employed to test the sight of the employees in the same way as most concerns have a doctor to look after the physical welfare of their employees. Fifth, to find out from wearers of glasses whether they have been influenced by advertising to ask by name for any particular kind of frame, mounting or lens. Sixth, to find out the percentage of people who have their eyes examined by an oculist and then take the prescription to an optician, and the percentage of people who have their eyes examined by an optometrist, who both examines the eyes and furnishes glasses. Seventh, to ascertain the percentage of people who are imprudent

Digitized by Google

enough to take a chance on their eyesight by buying glasses from the five-and-ten cent stores, fakirs, etc., without having their eyes examined by some one competent to do the work scientifically. To find out from the oculists, optometrists and opticians what trade-papers and scientific magazines they read, so as to make a tabulation by percentages of those which are most popular and the leading features of these magazines. Eighth, to secure information as to how much advertising is done by the oculists, optometrists and opticians, and the division of expenditure according to the kinds used and the usefulness of each kind. Ninth, to find out the source of supply of the oculists, optometrists and opticians, so that we may secure percentages as to which jobbers are the most popular and the reasons for their popularity. Tenth, investigation to be made with the optical jobbers to find out territories covered, selling methods pursued, advertising, etc.

USES OF STATISTICS BY ADVERTISING MEDIA.

All the instances thus far cited referred to uses of statistical data and methods by advertisers. They constitute one great unit in the advertising world. A second unit in the field of advertising is embraced in the term "media," that is, magazines, newspapers, trade journals, posters, etc. Advertising media utilize statistics most extensively for both administrative and research purposes. Let us consider some specific examples.

Circulation Statements. It is obvious that before an advertiser will sign a contract to use advertising space in a newspaper or magazine, that he will require a statement regarding the quantity and quality of the circulation. This is particularly important in view of the advertising rates now prevailing —a double page spread in a certain weekly costs $12,000 for one issue. Obviously, the advertiser wishes to study this information as to circulation, and to make comparisons of one magazine or newspaper with another. This was a difficult task as long as each magazine and newspaper would make up its own statement in the way it saw fit. For example, some newspapers would include in their circulation all the copies printed, while the advertiser is interested only in the number of copies sold. Then again, the advertiser wants to know how much of the circulation of a newspaper is in the city, and how

Digitized by Google

much is distributed in nearby suburbs. Other data of this kind were necessary for valid comparisons.

Of much interest in this connection was, therefore, the organization of the Audit Bureau of Circulations which it is worth our while to describe in some detail, because its reports are tabular, statistical statements, pure and simple.

The Audit Bureau of Circulations issues uniform, standard, and audited circulation statements of those publications that are members to advertisers and advertising agencies that are members of the Bureau. Inasmuch as different data are required for different classes of publications, five forms are used, one for each of the following classes: daily and Sunday newspapers; weekly, semi-weekly, and tri-weekly newspapers; magazines; farm papers; and trade, technical, and class journals.

Each of these forms contains very comprehensive statistical data regarding the publication concerned. The magazine form, for example, gives both the average paid and unpaid circulation, and separates the paid into mail subscribers and sales through newsdealers. Lest a certain issue be quoted by the publisher, that because of a special feature had an inflated circulation, the publisher must show the variation in the circulation of the last five issues. Then there is a tabular presentation of the distribution of the circulation by individual states and by geographic groups, both as to mail subscribers and sales through newsdealers. In addition, there is a statement of the circulation in cities of 100,000 and over, in cities of from 50,000 to 100,000 population, from 25,000 to 50,000, from 10,000 to 25,000, from 2,500 to 10,000, and under 2,500. This enables the advertiser to know what percentage of the circulation is in large cities and what percentage reaches the rural communities, a most important item.

The publisher is also required to specify the particular class, industry, or field covered by his magazine, also to give such information as the percentage of renewed subscriptions, data as to subscribers in arrears, whether premiums are offered to subscribers, etc. In short, there is a complete analysis of the quantity and quality of the circulation. The trade journal form goes even further in stating how many subscribers are

Digitized by Google

owners, managers, wholesalers or retailers, and in giving similar pertinent information.

Analysis of Advertising and Reading Matter. Statistical methods have been used very extensively for purposes of analyzing advertising and reading matter. Thus, the May, 1916, issue of *Advertising and Selling* contains an article on "A Graphic Study of Six Years Magazine Advertising." This article describes a study based on the number of lines of advertising carried by 60 representative magazines from January, 1910, to March, 1916.

The Butterick Publishing Co.—publishers of the *Delineator*, *The Designer*, and *The Woman's Magazine*—ran two advertisements in *Printers' Ink*, January, 1916. One gave an analysis of the reading matter that appeared in these publications during 1915. All the reading matter was divided into ten classes, for example, the home, kitchen and table, fashion, etc. The number of agate lines devoted to each was stated, and the matter was presented graphically in circular form, segments representing the percentages of lines devoted to each topic. The second advertisement presented a similar analysis of the advertising carried by that company in 1915. Going even further, the Curtis Publishing Co. has a checking department, "which regularly examines all the leading publications and measures and records the amount of space used by each advertiser. From these figures are prepared totals by weeks, months and years, by individual advertisers and groups, by large-space users and small-space users, by types of publications, and by other classifications."

The *Chicago Tribune* published a comprehensive analysis of Chicago newspaper advertising, containing in both tabular and graphic form complete advertising records of all Chicago newspapers for 1915. Display and classified advertising are treated separately. Each important industry or line of business is given in considerable detail. For example, the section on Automobiles and Accessories gives the number of agate lines of this class of advertising carried in 1915 by each of seven newspapers, and gain or loss over the previous year. Then the figures for each newspaper are given separately as to gas cars, electric cars and accessories. There is also given an

Digitized by Google

interesting table, based on a poll taken by the Firestone Tire and Rubber Co., of the rate per line per auto owner for each newspaper. Other classifications, such as Department Stores, Groceries and Food Products, are similarly treated.

At the 1916 convention of the Associated Advertising Clubs of the World, Paul E. Faust, touching upon this general topic in an address on "How Newspapers Can Organize Their Market," said:

> If a quarter of the income goes to wearing apparel products, it is evident that this is of great interest to the home. If 50 per cent. of the average income goes for foods, then food products can be seen to be a very live issue. If 10 per cent. of the income goes for furniture and house furnishings, it is reasonable to believe that these items are entitled to editorial attention so that the reader may be guided, counseled and given very much desired latest information on the subject.

He then quoted the figures of the *Chicago Tribune* on the amount of text devoted to various departments.

Realizing the importance of furnishing present and prospective advertisers with accurate data along definite lines of interest, The Curtis Publishing Co. instituted a Commercial Research Division, which has made a number of investigations of extreme interest. As to the functions of this Division, let me quote from *Obiter Dicta*, a Curtis house-organ:

> Broadly speaking, the duty of this division is to ascertain, by personal investigation, the conditions, manufacturing and selling, that exist in various trades and industries, and to interpret these signs into statistical expressions of tendencies and future possibilities.
>
> With this material in hand the Company's representatives may come to a manufacturer with something to give him that will be worth his serious attention—information as to the condition and extent of his national markets, facts that may suggest new merchandising methods, and tables and charts showing in graphic detail many points that before may have been matters for conjecture alone.

An instance of the work of the Curtis Research Division is an investigation of the textile field. In making this investigation, Mr. Parlin, the head of the division, visited 165 cities, conducting over 1,000 interviews, traveling over 32,000 miles. A year was spent in studying the textile industry, and the final

Digitized by Google

report consisted of four volumes—about 2,500 pages including over 100 charts and 12 maps.

Another interesting volume prepared by the Commercial Research Division is "An Encyclopedia of Cities" containing an estimate of the "trading" population and of the department store business in every city of the United States of over 5,000 population.

The encyclopedia is divided into states, with the cities in that state alphabetically arranged, and after each city is given the following information:

(1) The census population
(2) The native white population
(3) The trading population (that is the census population plus the number of people living outside the city who *regularly* trade in the city, and minus the number of people living in the city who *regularly* trade outside the city
(4) The predominating nationality
(5) The leading dry goods stores and their total estimated volume of business.

A most stupendous investigation of the automobile industry was also made by this Division. A four volume, 3,000 page report was the result. Among other reports of this Division are those on Food Products and Household Supplies, and Advertising Possibilities of Pacific Coast Industries. It is of interest to note that besides this Division for purposes of pure research the advertising department of the Curtis Publishing Co. also maintains a Statistical Division.

A case of similar research work conducted by a local newspaper is that of the *Chicago Tribune.* This newspaper, realizing that an out of town firm advertising or planning to advertise in the city of Chicago, would find it helpful to have definite data of local conditions, made a most thorough investigation of the city from various standpoints of business interest. This investigation is summarized in a book entitled "Winning a Great Market on Facts," published in 1916. A few paragraphs from this book will indicate the detailed use of statistical methods:

It took work—and lots of it—time and money for the *Chicago Tribune's* Merchandising Service Department to break up the population of Chicago into so many separate units, to

Digitized by Google

study, analyze and classify these districts, to be able to say with positive assurance "We know" and be conscious that the information was not based on theory or casual investigation, but on block by block canvasses. . . . .

It is for these reasons, then, that the *Tribune's* trained investigators, statisticians and merchandising experts divided this great city by the lake into 48 separate districts.

The *Chicago Tribune* through this Merchandising Service Department can give you the whole story of each of these districts, tell you the average rents, the buying habits of the tenants and how many Germans, Italians, Poles, etc., are in each district. In fact, put its finger on the hundred and one items of knowledge that determine why people buy and why they don't. . . . .

By this we do not mean to imply that the *Chicago Tribune* has made a personal investigation of the buying and reading habits of each and every housewife in Chicago. But we have expended a fund of time and money in investigating a sufficient number of housewives in each unit of Chicago's population, so that the *Tribune* can furnish you facts and figures, based upon the law of averages, which have been proved correct time and time again.

The *Tribune* has analyzed the Chicago market even more exhaustively than it has Chicago's population. The number of dealers in each line and their locations by district (that is, exact information as to the character of their trade), as well as the attitude of the dealer, small and large, and department stores toward advertised products, new lines, established lines, window and inside display, demonstration, and other means of pushing sales—all this is part and parcel of the *Tribune's* fund of data.

Newspapers and general magazines are not the only media that advantageously utilize the services of statistical data and research. The same tendency can also be noted in other classes of advertising media. Thus, a poster advertising concern—the Ivan B. Nordhem Co.—in an advertisement in *Printers' Ink*, May 4, 1916, refers to its statistical department that furnishes valuable information for guidance in the selection of territory. A book issued by the New York Car Advertising Co. contains considerable data on the various boroughs of New York City and on the individual car lines. Trade journals also gather and publish data on the industries that they cover. Moreover, advertisers and advertising agencies are expecting trade journals to be of service along these lines.

Digitized by Google

Publishers of farm journals form another group that are enlisting the services of statistics. Thus, the "Standard Farm Paper Year Book for 1916," issued by the Research and Service Department of the company publishing the *Standard Farm Papers*, contains the following introductory statement:

In compiling the Standard Farm Paper Year Book for 1916, we have endeavored to give the buyer of advertising space a comprehensive collection of agricultural information in condensed form.

We are headquarters for reliable, detailed information covering the production, resources and purchasing power of the American farmer. Should you desire specific information on any special phase of the subject we will be glad to furnish data covering your individual requirements without obligation.

The book then quotes, extensively, from statistics compiled by the Census Bureau, the United States Department of Agriculture, and other sources.

Just one more instance of the use of statistics by farm journals. In February, 1914, the publisher of *Successful Farming* made formal announcement of the establishment of a Consultation Department. The following paragraph appeared in the announcement, showing that the new department was really to be a statistical or research bureau:

For many years we had hoped that some day we could increase the line of detailed and comprehensive information covering the resources, production and purchasing power of our field which we are furnishing our patrons, and give more data on the possibilities of merchandise development through the different channels that reach the farm home.

A most interesting accomplishment of this department is the set of so called Definite Data Maps. In the words of the publisher, these maps "comprise a thorough analysis of the buying power of the American farmer and the best methods of distributing goods to him. Compiled by counties they give information for any part of any state of the Union. And geographically, portray at a glance the general centers of production." At the time of this writing, sixty Definite Data Maps were ready for distribution, including the following general heads: Soil Elements, Farming Area, General Valuations, Crop Yields, Farm Livestock, Operation Expenses,

Digitized by Google

Farm Tenure, Merchandise Distribution, and Farm Paper Circulation.

As an illustration of the practical value of these maps, a circular published by *Successful Farming* gives the following suggestion for their use to the mail-order advertiser:

> They will show him how he can most economically reach the different areas from which he can expect the most sales, where are found the best crop conditions, where the most money on deposit, where fewest retail dealers; as well as other essential things.

USES OF STATISTICS BY ADVERTISING AGENCIES.

The advertising agency constitutes a third unit in the field of advertising. The modern advertising agency is equipped to render the advertiser a very comprehensive and diversified service. It will plan and lay out the advertising campaign, conduct preliminary market and trade investigations, devise trade marks and packages, write the copy, attend to the art and printing, write booklets, select media, check results,—in short, an advertising agency is in a position to execute virtually every detail involved in the advertising and marketing of a product. Advertising agencies also utilize the services of statistics. In the first place, any of the instances of the uses of statistics by advertisers, described above, may be worked out by the agency in its service to the client. Then again, the "space buyers" of agencies make ample use of circulation figures in selecting the media for an advertising campaign. In addition, some agencies have established statistical or research departments for purposes of making market and trade investigations or other analyses that may arise in the course of their work. The following quotations present actual instances of the uses of statistical data and methods by advertising agencies.

In answer to a letter asking the editor of *Printers' Ink* what data agencies usually require in planning an advertising campaign, the editor states, in part:

> In many cases the advertiser is unable to furnish all of the information wanted. On this account, of late years, it has become common for agencies to conduct extensive trade and

Digitized by Google

consumer investigations, sometimes costing thousands of dollars and taking months of time.

N. W. Ayer & Son, a very large advertising agency, makes the following statement in one of its advertisements:

Our Plans Department makes an analysis and test of every product we advertise. It thoroughly digests all information relative to its manufacture and distribution.

In this department we have unusually complete and reliable statistics for visualizing distribution of population and wealth in relation to distribution and selling of merchandise. Moreover, no trouble is considered too great to get additional facts for the intelligent building up of an advertising plan.

Joseph H. Finn, President of the Nichols-Finn Advertising Co., states in an editorial in the *Chicago Herald:*

Modern merchandising campaigns are often the result of months of preliminary work. The heads of the business, the sales manager, the advertising manager and the advertising agent give to these conferences the best that each has. The entire merchandising procedure is carefully laid out, the market is studied, the demand is analyzed. Accurate information is secured on per capita consumption: competition is weighed with full allowance for its strength; inquiry is made into the relative prosperity of different territories.

And finally, let me refer to the Mahin Advertising Data Book, which contains, as stated on its title page, "facts and statistics in condensed form for the man who has advertising and merchandising problems to solve." This book of 500 pages, consisting almost entirely of detailed statistics, summarizes the uses of statistics to the advertiser as follows:

Statistics can show the advertiser where to find the people he wants to reach.

Statistics can help the advertiser in selecting his advertising mediums.

Statistics can help the advertiser in testing to what extent he is actually reaching the people he is after.

Report on Consumer Demand and Retail Conditions. Most extensive in scope and detail, and of unique importance to the manufacturer, the retailer, credit man, banker, advertising, and sales manager is the second annual report on consumer demand, the data for which was secured and compiled under the direction of Mac Martin, Chairman of the Educational

Digitized by Google

Research Committee of the Associated Advertising Clubs of the World. The report was published in the January, 1916, issue of *Associated Advertising* and comprises a stupendous piece of statistical compilation and presentation. A few paragraphs showing the nature and scope of the investigation, will be of interest:

The committee has not attempted to express opinions. It has sought only to gather facts and to classify and present them in a form which would make them of the most use to business men.

To obtain this information we have gone to the retail dealers in the lines closest to the consuming public—department stores, groceries, drug, hardware, jewelry and clothing stores. We have also obtained reports of stock carried as compared with the same period last year. We have asked them how their collections stand, so you may know the tendencies of the average consumer in regard to prompt payment in the United States and Canada at the present time. We have asked these dealers to report in definite percentages the increase or decrease in the amount they are spending for their own advertising. Every deduction and every percentage has been triple checked. . . .

Our able assistants here in Minneapolis have given night and day . . . to steady work in an effort to present this information fresh from the books of the dealers from the Atlantic to the Pacific, and from the Gulf to Hudson Bay, in the shortest possible time.

The tables, graphs, and maps that are a part of this report contain a wealth of statistical material of great value to all in any way interested in the marketing or merchandising of any commodity.

Uses of Graphs. In going over the material on which this paper is based, the writer was impressed with the extensive use that is being made of the graphic method of presenting facts. The business executive has neither the time nor the inclination to ponder over tables of figures in order to digest them properly. He would much rather have the facts presented to him in a more intelligible manner—in pictorial or graphic form. As C. C. Casey puts it in an article entitled "Graphic Sales Pictures That Analyze the Business" in *Printers' Ink*, January 20, 1916:

Digitized by Google

Sales and advertising statistics should be "read" before they get to the executive. They should not be *statistics* at all, but *pictures.* The tabulator should be taught to give the executive the *facts* which *he* ordinarily expects to get out of the figures himself.

Says W. H. Ingersoll, quoted in *Printers' Ink*, April 15, 1916:

Graphic charts may be devised for any character of data and show in visualized form important relationships which cannot be gotten in any other way, nor can so much information be gotten in such limited space by any other method.

The vital statistics of a business or a department, like the vital statistics of a nation or a government, may best be concentrated for preservation in the form of graphic charts.

Numerous other quotations could be given to show the significance of graphs to business executives. As for specific instances, the graphic method was used extensively in the investigations of the Commercial Research Division of the Curtis Publishing Co., and in the *Chicago Tribune's* Survey of the city of Chicago. The Definite Data Maps, described above, are graphs, pure and simple. The report on consumer demand also has a very large number of maps and graphs, and besides there is the suggestion that

Hundreds of other maps and charts can be compiled from these figures. . . . For example, the increase in population, the weather conditions, bank clearings, building permits, saving bank deposits, war orders—all these may be charted in areas and compared with consumer demand, to the benefit of those whose special lines indicate such additional charts.

Graphs are being used to plot territorial sales and advertising, to correlate advertising expenditures with sales results, to show increases or decreases in business, to compare the results of different departments of a business, to trace sales during various months of the year in order to locate seasonal tendencies, and to analyze other significant conditions of a business. The writer has even observed the use of graphs in advertisements to the consumer. For example, a recent advertisement contained a graph plotting the percentage increases in the prices of certain materials during 1914, 1915, and part of 1916.

Digitized by Google

A most interesting adaptation of the graphic method by an advertising agency is the Hanff-Metzger "Maplan," a merchandising blueprint. The "Maplan" presents, in graphic form, the interrelations of all the elements and efforts involved in the merchandising and advertising of a firm or product. An idea of the scope and detail of one of these blueprints can be had from the fact that they have been made as large as fourteen feet in length.

In concluding this paper, it is well to repeat that the writer set out to present a number of *actual instances* in which statistical data or methods have been and are being employed to help in the solution of advertising, marketing, and merchandising problems. In order to do this most effectively, it was considered best to *quote* those who were availing themselves of statistical aid.

After going through all the material, the writer was considerably surprised at the great extent to which statistics were contributing to advertising and merchandising efficiency, even though some firms were only unconsciously utilizing statistical methods. And yet, it is the writer's most earnest conviction that statistics can and should be consciously employed by more firms, and more extensively by many firms already taking some advantage of this aid in solving effectively and accurately their advertising and marketing problems.

Digitized by Google

REVIEWS AND NOTES.

Storage, Handling and Sale of Food in Philadelphia. By Janice S. Reed Lit.

Food Inspection Service in Philadelphia. By Bureau of Municipal Research.

The above two reports were published together by the Henry Phipps Institute of Philadelphia.

The report on the storage, handling, and sale of food contains the results of a very extensive and intensive investigation of conditions in Philadelphia, covering a wide field and including almost every variety of food dispensing store in the city.

One is not quite certain in reading this report whether it is intended for general propaganda or a scientific study of the subject. Should it be considered as a scientific study, it undoubtedly lacks the scientific evidence as to the value of certain food protecting measures and is completely devoid of chemical and bacteriological analyses which would show possibilities for contamination, stages of decay, and the effect that these conditions might have upon the consuming public. Considering the bulletin from the angle of a popular document, the first striking difficulty is encountered in the method of scoring and tabulation of the facts, which, while not complex in itself, is described in a manner that could hardly be understood by the lay mind without much effort.

On page 16, after presenting facts on the distribution by nationalities of store keepers, there is a statement regarding exhibits which seems to have no relation whatever to the preceding text. Here and there, there are slips of expression, such as "built stores" which are not defined in any part of the report.

In interpreting the figures contained in the tables, there is no effort made to fix upon the condition that would seem most striking and requiring the most concentrated effort towards improvement. In analyzing the scores, the discussion is so complex as to make it almost impossible for the non-statistical mind to follow the thought.

One is at a loss to understand the significance of such sentences as this: "Meat also is often similarly exposed and, worse still, offered for sale, side by side with live poultry." There are some reasons why live poultry should not be sold side by side with meat, but the reader not familiar with the technicalities of the subject is at a loss to understand why such objections should exist.

For the size of the report, the amount of tabular material is wholly out of proportion and the statistics are so complex as to require the most careful study on the part of the reader in order to derive any sort of conception as to their significance. Had the material been translated into a few small charts, showing the frequency of good and bad conditions in the various

Digitized by Google

stores, the bulk of the statistical data included in the report could have been reduced to the advantage of the reader.

A number of maps are appended which purport to show the distribution of various kinds of stores in the sections studied. It is with the utmost difficulty that any one can read the symbols inserted near the little circles indicating the various stores. It would have been very much to the advantage of the reader if some statement as to the actual number of the various kinds of stores had been placed on a map, instead of attempting to show them by symbols near the circles which show their location.

The whole report suffers from a lack of focussing of the information gathered and leaves the reader with no definite conception as to what the actual situation in Philadelphia is.

Had the method of presentation contained in the report on the food inspection system in Philadelphia been used in the report on the storage, handling, and sale of foods, the report would have been greatly improved and its value to the public at large vastly increased.

Regarding the section on food inspection service in Philadelphia, one is prompted to commend its clearness, and, as the departments of the city had carefully gone over the material, it is undoubtedly accurate as to facts.

In his introduction to this pamphlet, Dr. H. R. M. Landis speaks of *trichina spiralis*, which is found parasitic in hogs, and suggests the necessity for avoiding the consumption of hogs affected with this disease. We are wondering whether Dr. Landis is aware of the fact that the Bureau of Animal Industry of the United States has estimated that even a superficial inspection of all hogs slaughtered in this country with a view to eliminating those affected by the disease would run into the millions, and that even then the results would be questionable. The only alternative is a thorough cooking of the meat.

Dr. Landis also suggests that animals affected with tuberculosis be rejected for food purposes. Does Dr. Landis disagree with the United States Bureau of Animal Industry, which allows the use of such animals where the lesions are well-localized and the affected parts could be removed.

CAROL ARONOVICI,
Philadelphia.

BOARDS OF TRADE AND BUSINESS STATISTICS.

The yearbook of the Board of Trade of the City of Newark (N. J.) for 1916 contains an extensive "statistical summary of Newark," compiled under the direction of Secretary James M. Reilly. This feature of the annual report has served for many years to inform inquirers on the industrial and commercial resources of Greater Newark. It gives in short compass the essential facts of population, industry, and municipal and state government.

This organization has taken considerable interest in advancing the cause of industrial and commercial statistics. In 1915, it coöperated with the

Digitized by Google

Bureau of the Census in taking the manufactures census and proposes during the winter of 1916–1917 to devote one meeting to a consideration of statistics as an adjunct to business administration.

E. W. KOPF.

The Mathematical Theory of Probabilities and Its Application to Frequency Curves and Statistical Methods. By Arne Fisher. Volume I. Mathematical Probabilities and Homograde Statistics. New York: The Macmillan Company, 1915. Pp. xx+171.

Statisticians will welcome the appearance of Fisher's treatise because of the scarcity of such texts in the English language, and particularly because of the clearness and judgment with which the present book is written. They will await the appearance of the second volume with impatience; for it is in that volume that the more advanced development of statistical methods will be given. The present part deals almost entirely with the mathematics and the philosophy which underlie the theory of probability and which cannot be hastily put aside if a really fundamental knowledge of its applications to statistics is to be acquired.

The titles of the different chapters, with their initial page numbers, will indicate briefly the contents of the book:

I. Introduction: General Principles and Philosophical Aspects, p. 1. II. Historical and Bibliographical Notes, p. 11. III. The Mathematical Theory of Probabilities, p. 17. IV. The Addition and Multiplication Theorems in Probabilities, p. 26. V. Mathematical Expectation, p. 49. VI. Probability a Posteriori, p. 54. VII. The Law of Large Numbers, p. 82. VIII. Introductory Formulas from the Infinitesimal Calculus, p. 90. IX. Law of Large Numbers: Mathematical Deduction, p. 96. X. The Theory of Dispersion and the Criterions of Lexis and Charlier, p. 117. XI. Application to Games of Chance and Statistical Problems, p. 127. XII. Continuation of the Application of the Theory of Probabilities to Homograde Statistical Series, p. 146.

There has been so much discussion, if not controversy, concerning the foundation of the theory of probabilities, whether it should be upon an *a priori* or *a posteriori* basis that the author's careful discussion of the value and of the necessity of both bases is particularly welcome; he says in particular how careful one must be in the application of Bayes's Rule, and how a careless application of the Rule has been responsible for some of the distrust of the whole theory of *a posteriori* probability. Indeed it may well be said that the necessity for critical care in all statistical work is the impression which the reader gets most strongly from the text, as a quotation from the final section will show:

"A statistical research may be likened to the navigation of a difficult waterway, full of hidden rocks and skerries out of sight to the navigator. The amateur statistician, sailing the ocean in a blind and happy-go-lucky manner, often comes to grief on those rocks and suffers a total shipwreck.

Digitized by Google

The skillful navigator, the mathematically trained statistician, is always on the lookout for the sea marks. In the Lexian Ratio and the Charlier Coefficient of Disturbancy he recognizes a beacon light, often signalling 'Danger ahead.' He stops his engines. In case he does not possess the particular charts giving the exact location of the hidden reefs his prudence advises him to call a pilot to bring his ship safely into harbor. On the other hand, if he has reliable charts and knows his profession thoroughly he may venture forth and do his own piloting, by a study of the charts. It is to the study of such charts—*i. e.*, a special study of the higher statistical characterisitcs—that we shall turn our attention in the second part of this treatise. The reader who has followed us up to this point may perhaps feel discouraged by realizing how little he has gained in knowledge after having learned a mass of technical detail and formulas. We can quite appreciate and understand this feeling. So far, he has perhaps chiefly been impressed by the treacherous and misleading character of statistical mass phenomena, but to recognize a danger signal and thus avoid the pitfalls is one of the fundamental essentials in safe navigation in statistical research."

The typography of the book is excellent and the style also generally idiomatic, though occasionally there is evidence that the author wrote originally in Danish, and that in the translation an occasional slip has been made which will interest and not confuse the reader.

EDWIN BIDWELL WILSON.

Massachusetts Institute of Technology.

THE MEMORIAL VOLUME.

The delay in publishing the *Memorial Volume* of the American Statistical Association requires some explanation.

Prior to the outbreak of the Great War a number of articles on the History and Organization of Official Statistics in various countries had been received, but some of the most important essays were still missing. Those on Russian, German and Swedish Statistics have been finished *during the war*.

At the present time there are at hand and in page-proof articles on the official statistics of the following countries: Australia, Austria, Belgium, Canada, Denmark, Germany, Great Britain (including Ireland), Holland, Hungary, India, Norway, Russia, Sweden, and the United States.

Many difficulties have prevented the completion of the French, Italian, and Japanese studies. That a country so important statistically as well as in other respects as France, for instance, could not be omitted from our volume without seriously marring an important piece of work goes almost without saying. The alternative of printing these delayed articles in our regular QUARTERLY PUBLICATIONS did not appeal to the committee, and it has therefore been resolved to postpone publication until at least the promised study of French official statistics shall have been completed. And reason-

Digitized by Google

able assurance can now be given that this article will be finished within a few months as it has been possible to interest the French government in our plan. It is also confidently hoped that the delay will enable us to include an article on Japanese and probably one on Italian statistics.

We therefore reckon upon a volume in which practically all the important countries of the world will be represented through studies prepared for it by the foremost statisticians in each. Fortunately the delay does not age the material. It will be just as fresh and valuable as if it had been issued at the time first projected and worthy not only of our Association but of the science of statistics. It has required a large amount of labor, patience, and persistence to bring the volume up to its present state of completion. To forego the ideal set at the time the collection of the material was first undertaken merely for the sake of gaining a few months of time would certainly tend to spoil what we believe now promises to be a reasonably perfect piece of work.

As soon as practicable, members of the Association will be notified of the probable date of publication and of the conditions upon which it may be obtained.

JOHN KOREN,
For the Committee in Charge.

Digitized by Google

AMERICAN
STATISTICAL ASSOCIATION

NEW SERIES, No. 117. MARCH, 1917.

DO THE STATISTICS OF THE CONCENTRATION OF WEALTH IN THE UNITED STATES MEAN WHAT THEY ARE COMMONLY ASSUMED TO MEAN?*

By Allyn A. Young, *Cornell University.*

As I begin, I can imagine the critic asking: "But precisely *what* are the statistics of the concentration of wealth in the United States commonly assumed to mean? And where are these statistics and what are they?" Now although the title of this paper is not of my own phrasing, I am willing, if needs be, to defend it. We are a long way, of course, from knowing at all accurately just how property and incomes are distributed among the 21,000,000 or more families in this country. We do not even know whether the average family income is as low as $1200 or as high as $1600. We are not sure whether the share of the aggregate national income going to the richest 2 per cent. of our families is more nearly a fourth, or a fifth, or a sixth of the total. The poorest half of our families may possibly get a fourth or perhaps three tenths of the national dividend—just what the proportion may be is uncertain.

The picture is blurred, but its general outlines are nevertheless unmistakable. We are reasonably certain that such full knowledge as we might gain from a complete and accurate census of incomes would merely give precision and definiteness to our impressions. There would be, we are confident, no general upsetting of the notions we have already formed with respect to the state of things. Yet, this being so, the wonder is that most of us seem to accept the situation with a

*Paper presented at the joint meeting of the American Economic Association and the American Statistical Association, Columbus, Ohio, December 29, 1916.

Digitized by Google

large measure of complacency. If it is true that the richest 10 per cent. of our families get a third, or anything like a third, of the national dividend, and that the poorest fourth get incomes of less than $600, or at most, less than $700 per family, it might seem that we should not have to look further for an overwhelming indictment of our present economic system. Facts like these, some persons allege, must spell revolution.

Very likely we sometimes find it easy to push these troublesome statistics aside, saying to ourselves that viewing matters relatively, rather than absolutely, conditions are not so bad. Greater inequalities than these have existed in other times and other countries. Despite the relatively high average family income in the United States, incomes seem to be distributed here with rather less inequality than in some of the other great industrial nations. In Europe one must go to the relatively poorer nations, somewhere back of the first rank in industrial advancement, to find incomes that we can say with any certainty are apportioned more evenly than incomes in the United States. But this is evading the issue. If our economic life is diseased, it is small comfort to know that the disease is world-spread. And are we ready to admit our willingness to submit the account of our own social and economic state for comparison on even terms with the reports from older countries? Has much of our talk about democracy, equality of opportunity, and "a fair field and no favor" been idle? Is it for nothing that men from the old countries have built up a new civilization here, leaving behind them, as they thought, most of the handicaps of inherited privilege, of class distinctions, and of a long accumulation of economic advantages, unevenly distributed? Or is there something to be pondered in Mr. H. G. Wells' observation: "Mr. Britling could explain away the faults of England readily enough. . . . But there in America was the old race, without crown or church or international embarrassment, and it was still falling short of splendid"?

If wealth were unmistakably more concentrated here than elsewhere in the world, we should, very likely, be more alarmed about the situation than we now seem to be. But there are other things beside downright obtuseness and an easy satis-

Digitized by Google

faction with our standing relative to other nations that may affect our attitude. In particular, I suggest the sharp edge of these facts is blunted by the following considerations: (1) The picture they give is that of a cross-section of a changing, shifting current rather than of anything that might be interpreted as a definite and final outcome of forces now at work. (2) The meaning of inequality or concentration in the distribution of wealth and incomes is itself a matter of much uncertainty. (3) It is probable that complete and accurate statistics of the general sort that have already been used would still leave out of account some of the most pertinent facts. (4) Although we know that there is a high degree of inequality in the distribution of wealth, and although we know something of the general outlines of this distribution, the margin of uncertainty with respect to the facts is, nevertheless, so great that they have little compelling power. I shall discuss these points in order.

I.

One does not have to shut one's eyes to the gravity of the immediately pressing problems growing out of the unequal distribution of wealth in order to maintain that things are not so radically wrong as they may seem to be. One may retain a large measure of confidence in the fundamental wholesomeness of a competitively-ordered society, and yet concede the concrete injustice of existing conditions.

During the last one hundred and fifty years the production of wealth has grown faster than the population. And new productive methods, new kinds of wealth, new forms of business organization, have followed, one upon another, in rapid succession. There are few varieties of income which can be expected to keep pace, year after year, with this process of restless change. In a dynamic society more is produced than has to be imputed, in the form of earned income, to the efforts of the rank and file of the productive army. Economic progress yields what is, for the time being, a disposable surplus. The shares in this surplus are the principal stakes in the great game of business enterprise. Here are the roots of some of the most striking phenomena of the concentration of wealth, and here there are the widest disparities between service and

Digitized by Google

rewards. There are prizes here for efficiency, for alertness, for foresightedness, and for luck and for unscrupulousness as well. Here are the chief sources of the gains of monopoly, of financial manipulations, and of strategic competitive advantages. Here are found most of those capitalized income-yielding opportunities which we know under such names as franchise values, corporate excess, and good will. The fruits of progress are not at first apportioned equally among all the coöperating producers, nor do they go, in major part, to the pioneers of science and industry who have made the largest effective contributions to the knowledge that makes progress possible. They go, rather, to those who actively and successfully contend for them.

But there comes, out of it all, increased demands for labor, for savings, and for productive agents of all kinds. Unless the supply of some productive agent is increasing with undue rapidity nothing can prevent it, as the volume of output grows, from commanding a higher price in the market. Monopoly may crumble and other business advantages may be snatched away by competition, but the forces working toward the diffusion of the product operate relentlessly and surely. Every bit of ground gained by the rank and file is tenaciously held, and becomes a starting point for yet further progress.

This impressionistic picture of an economic process must, for present purposes, stand without the support of a detailed analysis. But it is, I believe, in essential harmony with the doctrines of most of the schools of economic theory. Its relation to the interpretation of statistics of the distribution of incomes and of property must be fairly apparent. Such statistics give a cross-section view of things; they speak for a given year or a given moment of time. But the operation of economic forces is not thus synchronized. The concentration of wealth that exists at any one time in a progressive society is in part an outcome of the fact that the productive army does not advance in even formation in its attacks upon the new sources of wealth. There is an advance guard and there are laggards and stragglers. No one knows just how far the concentration of wealth could be reduced, just how far the forces tending toward diffusion would operate effectively, if economic

Digitized by Google

progress should cease, if our dynamic society should crystallize into a static state. But that distribution would be much evener seems fairly certain. The concentration of wealth is in part one of the unwholesome fruits of progress. If concentration and diffusion are processes going on simultaneously, statistics show us merely the extent to which, at any one time, the forces making for concentration are ahead of the forces making for diffusion.

It may be objected that this is to no purpose, or at any rate beside the main point, that present inequalities in the distribution of wealth and the social problems growing out of them are none the less real. It is, of course, small comfort to those who are getting less than their fair share of the present product, to know that they or their successors are likely to get more, particularly if along with this assurance there goes the probability that new inequalities are to be built up about as fast as the old ones are leveled down.

Now it is perfectly true that to impute these inequalities to one cause rather than another does not lessen their gravity. But the right diagnosis of a disease does have an important bearing upon the choice of remedies. And so the right interpretation of the statistics of the concentration of wealth must strengthen the convictions of those of us who are opposed to anything like an arbitrary leveling down of fortunes or a revolutionary change in the general structure of our economic life. It must give us confidence that we are on the right road in our efforts to control monopoly, to restrict some of the abuses of corporation finance, to do away with unfair privileges and advantages, to establish and enforce fairer standards of competition, to strengthen the bargaining power of the weaker, to reduce the extent to which accrued inequalities are transmitted from one generation to another by inheritance, and, by controlling immigration, to check the tendency of the process of diffusion to become a process of dilution. It may well be that this road is longer and has more turnings than we now imagine. But there is nothing in what we now know about the distribution of wealth to suggest that this road—the road toward equality of competitive opportunity—is not the right one.

Digitized by Google

II.

I pass to a more technical question. What is the precise meaning of the concentration of wealth? By what standards shall we measure it? In general, I think, statisticians have been accustomed to use "concentration of wealth" and "inequality in the distribution of wealth" as loosely interchangeable terms. Now any departure from perfect equality in the distribution of wealth means *inequality*. But is *concentration* to be defined so broadly as this? Wealth might be distributed unequally, without there being any amassing or concentration of any relatively large part of it in the hands of any one group or portion of society. Concentration means, then, a particular kind of inequality in distribution. And while, statistically speaking, any perceptible degree of centralization must be deemed concentration, yet the social problem of the "concentration of wealth" is very certainly the problem of its undue or excessive concentration. But we have no definite standard of what constitutes justifiable, permissible, or normal concentration. And so the statistics are made to indicate merely the gross departure from a condition of absolute equality in distribution. One has to be on one's guard, therefore, against imputing to them a significance which possibly they may not have.

There ought to be agreement, it would seem, about what constitutes *equality* in the distribution of wealth. Pareto's interpretation of the meaning of his own well known index of inequality in the distribution of incomes is based on the expressed assumption that when the number of persons with incomes of less than a given size increases relatively to the number of persons with higher incomes, the inequality of income distribution diminishes. I do not quarrel, as some have, with this definition of relative equality and relative inequality in distribution. Given a definite income range, with small incomes more numerous than large incomes, Pareto's hypothesis is in general consistent with the common notion of the meaning of equality in distribution. But, so far as I know, it has not been noticed even by his critics that Pareto, brilliant mathematician that he is, made a curious slip in interpreting the relation of his index to his definition of inequality. His

Digitized by Google

index does not increase, as he supposed, with what he deemed inequality in the distribution of incomes, but decreases. It can be used, as Benini and Bresciani have used it, in this inverted fashion. At best, however, it is a poor index, being merely a rough measure of the evenness with which income receivers are distributed through the income range. It gives no simple and definite standard of comparison. The index becomes unity, for example, when the relative numbers of the persons receiving incomes of stated sizes are inversely proportional to the squares of their incomes. How far, in general, the notion of an even distribution of income receivers throughout an income range differs from the common notion of equality in income distribution may be inferred from the fact that, with an absolutely even distribution of income receivers over any income range, the richest fourth of the population would get seven sixteenths of the total income, while the poorest fourth would get but one sixteenth.

For the most part, however, equality of distribution is interpreted literally; that is, it is taken to mean absolute uniformity in the distribution of income. Thus, when a statistician throws his estimates into the familiar form that assigns a certain (large) proportion of the aggregate income to a certain (small) proportion of families, the comparison inevitably implied is with a state of things in which 50 per cent. of the families get exactly 50 per cent. of the aggregate income and 10 per cent. of the families get 10 per cent. of the income. And so with Dr. Lorenz's graphic device for representing the way in which such proportions depart from the line of absolutely equal distribution. So, too, with the index of concentration which Professor Corrado Gini has suggested as a substitute for Pareto's, but which increases when Pareto's decreases, and which becomes unity when one income receiver gets just as much income as another. So also with Gini's other index, which takes into account the sum of the differences between each income and every other income. This index, it is interesting to note, may be interpreted as a summary arithmetical expression of the degree of concentration denoted by the Lorenz graph. So, finally, with all of those measures, such as the average deviation, probable error, and standard deviation, which indicate in a

Digitized by Google

general way the extent to which incomes as a whole differ from the average income. The significant thing is that all of these ways of expressing the degree of inequality in the distribution of wealth use as a standard or reference of comparison an absolutely equal and uniform distribution.

Some or all of these measures are useful in comparing the distribution of wealth in different countries or at different periods. But none of them is of much help in forming a judgment with reference to the degree of undue or excessive concentration that may exist. The degree of departure from absolute equality, however measured or stated, must itself be referred, if not explicitly, then in some vague way, to a standard of normal or justifiable concentration. A dead level of uniformity is neither practicable nor desirable as an ideal of distributive justice.

A concrete example may give point to this consideration. Suppose that incomes in an imaginary society were distributed symmetrically around the modal or most common income, in the form of a normal frequency distribution. This might represent either one of two things: (1) A normal distribution of ability and a perfect proportioning of income to ability; (2) a random or chance distribution of incomes, under the influence of complex but unbiased forces. This second condition would be consistent with the existence of real equality of opportunity, broadly understood, coupled with the presence of a myriad of small circumstances that might deflect one towards a lower or a higher portion of the income range. Now suppose that the average family income is $1500 and that half of the families get incomes that are within $200 of this average. Under such conditions the richer half of the families would get 58 per cent. of the aggregate income and the poorer half would get 42 per cent. Increase the dispersion of the distribution somewhat, so that half of the incomes are between $1000 and $2000. Then 70 per cent. of the aggregate income would go to the richer half of the population, and 30 per cent. to the poorer half. Increase the limits between which half of the incomes fall to $800 and $2200, and the portion of the aggregate income assigned to the richer half of the population becomes 78 per cent., leaving 22 per cent. for the poorer half.

Digitized by Google

I do not think that Dr. King's recent estimates err in the direction of underestimating the present inequality in the distribution of incomes in the United States. He assigns about 27 per cent. of the aggregate income to the poorer half of the families and 73 per cent. to the richer half. But this is a slightly smaller degree of concentration than would be given by a normal frequency distribution, with half of the incomes falling between $900 and $2100. This suggests that no simple or general statement of the degree of concentration can give, by itself, an adequate notion of the extent to which the existing distribution of wealth has to be deemed unsatisfactory. And instead of tabulating the statistics in the misleading form of the proportions of aggregate income or property in the hands of stated proportions of the population, it is better to use a simple frequency distribution, showing the relative numbers of income receivers or property owners in the different income or property classes. Such frequency distributions can be adequately described and compared, one with another, and with various ideal schemes of distribution by the use of the constants devised by Pearson for measuring their spread, skewness, and curvature. Such a handling of income statistics serves to focus attention upon the really important things, which are the upper and lower limits of the income scale and the manner in which income receivers are distributed between these limits. The *amount* of concentration, the amount of departure of a condition of uniform incomes, does not matter so much as does the particular *form* of the income distribution underlying the concentration. An identical general degree of concentration may result from a fairly good and a very bad distribution of incomes.

The worst thing in the present situation is undoubtedly the extreme skewness of the income frequency curve. The mode —the most common income magnitude—is very close to the lower limit of the distribution. Then the income curve descends rapidly as the higher incomes classes are brought under review, reaching a condition of relative attenuation at incomes of only a few thousand dollars, but stretching on for an absurdly great distance before the maximum incomes are reached. The problem of poverty and the problem of great fortunes

Digitized by Google

are the problems of the lower and upper limits of this income curve. But, seen rightly, the problem of great fortunes is only part of the larger problem of the general skewness of the curve, the problem, that is, of the extremely small average differences in the incomes of persons in the lower part of the income range and the unduly rapid increase of these average differences as the view is shifted to successively higher income groups. Put concretely, that 10 per cent. of the families in the country get possibly three fifths or two thirds of the aggregate income ceases to appear principally as a problem of large fortunes, when it is realized that to include the richer 10 per cent. of the families, one has to go down to somewhere between the $1200 and $1800 income levels. The most serious aspect of the distribution of property and incomes in this and other countries is not the presence of a larger or smaller degree of "concentration," but the general distortion of the whole income scheme, reflecting as it undoubtedly does the presence of a high degree of inequality in the distribution of opportunity.

III.

I do not propose to discuss here the meaning and proper interpretation of all of the various available or possible sources of information respecting the distribution of property and incomes. Some of these matters, I understand, are to be dealt with by Dr. King in the paper which is to follow. Nor can I stop to attack the difficult general problems of the proper definition of wealth and of income for the purpose in hand. But I want to mention two points which seem to me to have an important bearing upon the degree of inequality in distribution shown by the statistics.

In the first place, what disposition shall be made in income statistics of the appreciation and depreciation of capital values? I do not refer here to such appreciation and depreciation as comes from improvements and betterments on the one hand and wear and tear on the other, but to the changing money value of unchanged things. These changing capital values, it is true, are in large measure a result of changes in the probable future income-yielding power of the things valued. But it does not follow that to count appreciation as income involves

Digitized by Google

double counting. If savings are to be counted as income, and I believe that they must be, for the purposes of income statistics, appreciation in the money value of income-yielding goods or rights, being closely akin to savings in a number of important respects, must also be counted as income. It is no part of the annual product, properly defined, but it is part of the annual *distribuendum*. It enters into the annual changes in the relative economic positions of the different members of society. Any one income receiver can convert his share, if he wishes, into money income and then into real income. It is impossible and, I think, not at all necessary to distinguish between such appreciation as merely records the shrinkage in the purchasing power of the dollar and such appreciation as comes in other ways. So long as appreciation, of whatever sort, is distributed in other manner than in proportion to the other elements in current incomes, it has to be taken into account in income statistics, or at least in the interpretation of the significance of such statistics.

The census estimate of national wealth for 1912 exceeds the similar estimate for 1900 by about $100,000,000,000. Some of this—I doubt that it is more than a third—represents an increase in the national inventory measured in terms of physical units, together with the increase in the unit value of short-lived equipment and products—the current assets of the nation. The balance of this increase consists very largely of net appreciation in the value of land and other durable forms of wealth. And there is an enormous amount in franchise values, monopoly rights, and good will and other vendible competitive advantages that does not enter into the statement. Altogether, it is likely that the recent average annual addition to the national *distribuendum* on account of net appreciation has amounted to as much as $5,000,000,000 and possibly as much as $7,000,000,000, making a difference of perhaps 20 or 25 per cent. in the national income statement.

One of the defects of British income statistics is that they do not cover gains from increases in capital values, even when these gains are realized in actual sales, except where speculation is carried on as a business. Now the effect of the inclusion of these items upon the apparent degree of inequality in the

Digitized by Google

distribution of incomes will depend very largely upon whether they are counted as income only when realized by actual sale or set down yearly as accruals. Take the case of a farmer who bought a 100-acre farm 10 years ago. Since then it has been increasing in money value, as many farm lands have, at the rate of $2 per acre per year. He now sells it with a gain of $2000. If all of this profit is set down as income for this year, it lifts him, for the once, into the higher income groups and increases the general skewness of the distribution of incomes for the year. But if accruals had been counted at the rate of $200 a year for 10 years the increase would have been unlikely to have carried him up above the average income for the country. In fact, the skewness of income distribution might very possibly have been reduced. Moreover, if there were 1,000,000 other farmers similarly conditioned, there would be the choice between counting 1,000,000 moderate gains or an average of perhaps 100,000 large gains annually. And so with those enormous increases in the values of business equities which have been a very important source of large fortunes, the method of accruals will increase the skewness of the income curve very much less than will the method of realized sales. For the purposes of income statistics there can be little doubt that the method of accruals should be preferred. To wait until the farmer's increased land values are realized in sales exaggerates the farmer's real position in the income scale in one year out of the 10 and understates it the other 9 years.

My other point relates to the statistics of the distribution of property. The most important facts that we have, and the ones most frequently quoted, have been gleaned from probate records. I confess that I have little confidence in the inferences that have been drawn from these figures. The principal difficulty is the difference between the age distribution of decedents and the age distribution of the living population. In England, for example, while the average age at death of that part of the population which lives to be at least 20 years old is over 60 years, the average age of living persons over 20 is about 40. Now as men grow older they often grow richer, and this fact has been shown by Mr. Bernard Mallett to have a very important bearing upon the proportion which the value of the

Digitized by Google

estates annually passing from one generation to another makes of the aggregate capital value of the national wealth. For example, in the fiscal year 1913–14, the average value of the taxable estates left by persons dying in that year was £4150. But those dying at between 25 and 35 years of age left estates of an average value of only £1600. The average values of the estates which had been accumulated by persons dying between the ages of 65 and 75 was nearly £4200, while persons who lived to be more than 75 left estates with an average value of over £6000. In the United States, where fewer persons start with the initial advantage of inherited wealth, these differences are likely to be considerably greater. Now the danger is not that the probate records will give us an exaggerated idea of the aggregate money value of property holdings, for we have other and better sources of information on this point. But as used by Dr. Spahr and even as used by so careful and competent a student as Dr. King, they give an impression of a higher degree of inequality in the distribution of property than really exists. Not only do many men grow richer as they grow older, but some men grow rich faster than others, while some men, especially among those with little property, grow poorer as they grow older. It follows that the inequality of possessions among persons at the end of life must be very much greater than among the living population. Statistics of the sizes of estates admitted to probate are nearly worthless unless they are accompanied by statistics of the ages of the decedents. And with the probate statistics thrown out of court, there remains virtually nothing that gives us any adequate notion of the distribution of accumulated wealth in the United States.

IV.

Without knowledge of the way in which the ownership of property is distributed, with our income statistics largely guesswork and their meaning and significance uncertain, we have naturally set the problem of the personal distribution of wealth in terms of the things we could see and be certain of: enormous fortunes at one extreme, a great mass of workers near the poverty line at the other. Right though we be in taking these things seriously, we should hardly be willing to rest content

Digitized by Google

with a social program made up of anti-monopoly laws, inheritance taxation, and labor legislation, and a philosophy of economic reform compounded of philanthropy, of an avowed faith in industrial democracy, and of an attitude of jealous aggression toward great riches. As economists we are interested, or ought to be interested, in knowing whether the present economic scheme is bound to end in disaster, whether, with a little patching, it can be made to get along, or whether its essential principles need to be safeguarded and maintained by giving a new meaning and a new emphasis to real equality of competitive opportunity. In no other way could so much light be thrown upon this fundamental problem as by a thoroughgoing and authoritative study of the actual distribution of property and incomes. We lack satisfactory methods of analysis and interpretation, but most of all we lack the facts. General impressions will not suffice; in matters like this we are right in deferring judgment until we know, and know accurately. It is not a matter of the extremes of the income scale, but of the whole range. In statistics the undistributed middle is prone to be quite as dangerous as in logic. For the study of actual present tendencies the analysis of middle-class incomes is fundamentally important. It is one thing to explain the causes of poverty or the causes of large fortunes: it is another thing to know just how sound our economic life is at its core.

The task is much too large for any individual scholar. The first step is to secure adequate sources of information. With income taxes just coming into importance as sources of federal and state revenues, is the time not ripe for the economists and statisticians of the country to make a statement as to precisely what information they want and why they want it?

Digitized by Google

DESIRABLE ADDITIONS TO STATISTICAL DATA ON WEALTH AND INCOME.*

By Willford I. King, Ph.D., *Assistant Professor of Statistics, University of Wisconsin.*

I.

The General Characteristics of Available Statistics. As soon as one essays the task of analyzing existing information concerning American wealth and income and their distribution, he is quickly faced by the fact that, in almost every instance, the data are strikingly incomplete. But this scantiness of desired figures does not come to pass because of any shortage in the effort which has been expended in the collection of statistics. My limited experience has convinced me that there is no subject under the sun concerning which facts have not been tabulated. Should you desire to know the number of grains of sand in the sea-bottom, the number of salmon caught in the Columbia, or the number of "suckers" caught by "blue-sky" corporations, the figures have been computed. If you are curious to ascertain how many days' labor it took to build the pyramids, how many germs you consumed at this morning's breakfast, or how many times per page Shakespeare used the word "and," rest assured that the data are awaiting your examination. But, as a prominent advertisement used to read, "It is not so much to know the law as to know where to find it," and most of this marvelously varied collection of statistical information is buried far beyond the depths to which catalogues, dictionaries, or bibliographies give the slightest clues.

There is, then, rather a superabundance of statistical, or pseudo-statistical, information reposing in our libraries, but, should we be interested in the problems connected with the wealth and income of the American people, we soon discover that the figures which we desire are not obtainable in any continuous or systematized form. And this may be said without

*Paper presented at the joint meeting of the American Economic Association and the American Statistical Association, Columbus, Ohio, December 29 1916.

Digitized by Google

in any way disparaging the able officials who have been in charge of the Census and our statistical bureaus. The funds at their disposal are distinctly limited and they can only collect data concerning facts for which they believe the demand is most general. In order to assure continuity, which is such an essential feature if statistics are to prove of value, it is imperative to follow rather closely in the footsteps of one's predecessors, and, hence, the nature of the material gathered is largely determined by the forms outlined many years ago to meet the demands of that day.

Besides, the chief of the Census Bureau or of the statistical office would need the gift of second sight to anticipate the thousand and one ideas which some one or another will attempt to elucidate or demonstrate by the help of the official figures. If a statistical problem of any intricacy is to be solved accurately it is nearly always imperative that the data must be gathered with this specific end in view. If, however, the government collects its figures for this particular purpose they are often too highly specialized to prove of any great utility in answering other questions. Is there any escape from this dilemma?

In so far as it is practicable, it appears that the solution of special problems should be left to private statisticians or organizations and that the government should continue to collect the raw material with which these private investigators begin their researches. But there are some specific problems which are too costly to be undertaken at individual expense and yet, at the same time, are of too great public importance to be left to private initiative. If light is to be thrown upon such questions it must be done principally through governmental assistance, and there appears to be no sufficient reason why this should not be forthcoming.

Our federal Census Bureau has properly devoted itself in most instances to the collection of information of a broad general nature. The other bureaus of the national and state governments sometimes have imitated the Census Bureau in this respect, but sometimes they have gone to the other extreme and have engaged in the most highly detailed investigations in some limited field. This procedure is to be com-

Digitized by Google

mended although it seems occasionally that some improvements might well be made in the selection of the special problems. Whenever it is deemed advisable to enter such narrow fields, it is desirable that the problems to be solved should be those in which large numbers of people are interested or those which vitally concern the national welfare.

The federal and state bureaus of labor present hundreds of volumes of figures concerning the wages of laborers, the extent of unemployment, and the prices of various articles, but should you attempt to find a continuous and comparative record in any governmental document showing whether the price per hour of labor is advancing faster for farm hands or for miners, for railway employees or for clerks in stores, you would seek in vain. Should you desire to know whether the price of house rents tends to keep pace with the price of food, information would prove most scanty. Should you hope to ascertain whether union men or non-union men were more frequently unemployed, you would be disappointed.

A tremendous amount of effort has been expended in studying questions of merely local or temporary interest and very little headway has been made in coördinating these scattered endeavors, uniting the results, and presenting them to the public in a simple and intelligible form. For example, the United States Bureau of Labor Statistics collects data on wages in the manufacturing field; so does the Census; so do the labor bureaus of some of the states. At the same time, it is most difficult to find any continuous records of the wages of miners, farm hands, clerks, domestic servants, and other very numerous and important classes of workers. In some instances, wages are given by the hour, in some by the day, in some by the week, and occasionally by the month or year; but only rarely are we informed as to the average number of hours worked in any of the longer periods, hence accurate comparison is manifestly impossible.

According to my, perhaps, archaic views, energy too frequently has been expended in trying to demonstrate perfectly obvious facts. We all know from common observation that many men receive such low wages that they cannot support their families in comfort. It is also perfectly evident that the

Digitized by Google

poorer families spend their *entire* income for something—and these facts are apparently the leading ones that have been developed by many of the recent budgetary studies. The difficulty seems to be that many specialized inquiries are undertaken without first clearly defining the problem to be solved and, in every such case, the usefulness of the results is likely to approach zero.

It is probably too iridescent a dream to ever attain reality, but it would certaiuly be most helpful if the collection of all statistics of a given type might be centralized in some one bureau and published in standardized form from month to month and year to year. Before beginning operations, this bureau should devote very much time and energy to ascertaining exactly what problems scientific men believed most in need of elucidation and, having formulated these problems in the most exact fashion, it would then be the work of the statisticians to devise ways and means of solving them. It would seem that the talent of both of the Associations here represented might very well be utilized in framing the new topics of investigation upon which they believe further knowledge is most vital and in devising ways and means for attaining this end.

I shall take the liberty now of suggesting certain fields of inquiry which seem fertile and yet inadequately tilled at the present time.

II.

Wealth and Its Distribution. One query to which it seems that a more exact answer is generally desired is: How is wealth distributed among American families and how is this distribution changing?

The Massachusetts studies of estates throw most interesting light upon this question. Studies from probate records are necessarily limited in scope because they show wealth only at the termination of life and because the estates of the greater share of decedents are never probated. The latter is, of course, the much more serious defect. But neither of these objections seriously affects the usefulness of such data as a gauge of *changes* in distribution. A quarter of a century has elapsed, however, since the last period reported upon by the

Digitized by Google

Massachusetts Bureau and it would certainly seem worth while to have the series of comparisons, so admirably begun, brought down to date. Similar investigations in New York and a few sample states in other sections of the country would doubtless add most valuable information.

The tax rolls are the basis of existing Census figures concerning the aggregate value of the wealth of the people of the United States. The attempt to adapt to this purpose the valuations set by assessors is always destined to meet with only partial success. Values for purposes of taxation are habitually underestimated. The plan of correcting this underestimation by dividing by some assumed ratio of assessed to true value will always be a crude method at best. In most communities, this ratio has never been worked out, and, in but few places, is it determined with even a reasonable approach to accuracy.

Another serious objection to using local tax reports as a basis of wealth estimation is that it is usually impossible to obtain therefrom any idea whatever of the distribution of property among families or individuals. Taxes are collected, not as a basis for statistical information, but in order to obtain funds, and, hence, the local or state authorities take little account of anything but totals. For these two reasons, therefore, it appears that any adequate statistics of existing wealth distribution must be obtained through the ordinary schedules of the Decennial Census. Errors from inquiries of this nature will doubtless be very numerous but they will tend to be compensating. Many men, because of habitual secrecy and reticence concerning their business affairs will report their assets far below normal. Others, however, will exaggerate in order to magnify their own importance. As to which instances will occur more frequently, it is hard to say but, at any rate, there will be a noticeable tendency for errors to cancel out. The same facts would be true as regards reported liabilities though the tendency in this case would presumably be stronger to under- than to over-estimation.

The first step in the Census inquiry would be the drafting of the best possible question schedule. Much careful study would be needed before the final forms were ready for the printer. The following outline is suggested as a rough tenta-

Digitized by Google

tive classification of the various categories of wealth belonging to each individual:

I. Assets.
  A. Indirect Natural Resources.
    1. Land.
      a. Farms.
      b. Urban business sites.
      c. Miscellaneous varieties.
    2. Mines.
    3. Miscellaneous.
  B. Transition or Capital Goods.
    1. Buildings.
    2. Machinery and equipment.
    3. Live stock.
    4. Raw materials.
    5. Dealers' stocks.
    6. Miscellaneous.
  C. Direct or Consumption Goods.
    1. Residence and park land.
    2. Houses, hotels, theatres, churches, etc.
    3. Furniture, clothing, jewelry, and miscellaneous.
  D. Claim Goods.
    1. Notes, mortgages, bonds, etc.
    2. Stocks.
    3. Insurance policies—cash surrender value.
    4. Miscellaneous.
II. Liabilities.
  A. Current accounts.
  B. Short term notes.
  C. Bonds and notes secured by mortgage.
  D. Miscellaneous.

Our inhabitants are already required to give much of this information so that the new questions to be added to the schedule would not be numerous. If it is feasible to learn the value of farm land, factories, etc., it likewise seems probable that similar knowledge can be gathered concerning holdings of stocks, bonds, and other securities.

Once the data were collected, the next problem would be their classification on a logical basis. To determine what is logical we must first decide upon the questions which it is most imperative to answer. As previously stated, such a decision as this would seem to be a legitimate function of such organizations as the American Statistical and Economic Associations

Digitized by Google

and should, of course, be arrived at only after most mature consideration. As topics at least worthy of mention, I would suggest studies of the relationship between the net assets of the family and

a. Size of family.
b. Age of head of family.
c. Race.
d. Nationality.

It may be well to devote a few words to the reasons for each of these classifications.

The first would throw light upon such questions as the following:—

1. Does great wealth mean larger or smaller families?
2. Does a large family necessarily lead to poverty?
3. How does wealth in general tend to vary with the size of the family?

The second classification would help in answering this query:—Do we have a normal tendency in the United States for a man to accumulate property as he grows older? If so, at what rate? Do all classes of the population participate in this tendency?

The third classification would be necessary in order to separate out the white race so as to obtain a fair basis for answering the questions just propounded. It would also be of interest in enabling us to compare the Caucasian and Negro races and, if the inquiry were continued decade after decade, the relative economic progress of whites and blacks might thus be measured with a reasonable degree of accuracy.

The fourth classification, that on the basis of nationality, is one which might be elaborated indefinitely, if funds were adequate, or, on the contrary, it might be confined merely to the comparison of property holdings of those persons who were native born of native fathers, who were native born of foreign fathers, and who were foreign born. This simpler analysis would probably answer the requirements of most investigators and would immensely reduce the extent of tabulation necessary.

Digitized by Google

Many other classifications and lines of inquiry would doubtless suggest themselves to nearly every economist, and the above are offered only as being among those which would, perhaps, have the widest appeal.

III.

The Aggregate National Income and Its Distribution. A study of wealth distribution is interesting because it deals with the rights of control. Property necessarily means economic power and is ordinarily the key to other kinds of power as well. A man's control over his own time, over his own freedom of movement from place to place, over the activities of other men, over the uses to which goods shall be put, in short over the whole modern system of production varies directly and almost proportionally with the value of the property rights which he possesses. To the ambitious man who believes that leadership is the only worthy goal or to that political scientist who considers governmental control the chief end of human progress, a study of property distribution seems most vital. On the contrary, to those who think more of immediate economic welfare, who are interested in the comforts and conveniences of everyday life, the question of wealth or property may seem of little import. Such persons will naturally turn their attention to the various types of income rather than to wealth.

The kind of income with which people are most vitally concerned is, of course, real or psychic income. Thus far, however, this particular species has successfully defied all attempts at accurate statistical measurement. Apparently, the commensurable phenomenon which most nearly resembles psychic income in the nature of its variations is income in purchasing power and this is merely a calculation of what book or money income would be if the price level never varied. Since, in reality, the price level is always fluctuating, we can only measure income in purchasing power by computing an average index of the prices of things which some designated class of persons actually purchases and then compare the average relative price changes in these commodities with the average relative changes in the book or money incomes of the persons compos-

Digitized by Google

ing the class in question. In short, we attempt to measure the supply of economic goods available for a given person or class by means of the roundabout device of introducing commensurable money values into both terms of a ratio with the knowledge that the monetary factor will cancel itself and disappear.

If, therefore, we desire to add anything worth while to our existing knowledge concerning income distribution or changes, we can only hope to succeed by improving our data concerning book or money incomes and various kinds of indices of prices.

In the United States, the only available source of information concerning income distribution among the wealthier classes is contained in the reports on the income tax. This is necessarily faulty because of the natural tendency to evade the burden of taxation. The federal law exempts incomes to such an extent as to make it throw light only on the receipts of the relatively wealthy. The Wisconsin law varies the exemption with the size of the family thus making it very difficult to utilize the data. By going back to the original records, Professor Henry M. Trumbower, now a member of the Wisconsin Railroad Commission, was able to regroup the cards in such a way as to show the reported incomes including exemptions. However, since married men with incomes of $1200 or under are not required by the Wisconsin law to file schedules, and since these comprise the majority of heads of families, it is manifestly impossible to arrive at any complete classification. It is not an entirely safe assumption anyway to suppose that Wisconsin fairly represents the other 47 states of the Union in income distribution, even though we know that it is an average state in many respects.

If we now turn to the problem of estimating income distribution among the poorer classes, we find here, likewise, but few accurate indicators. The investigations of the United States Bureau of Labor in 1903, the study by the British Board of Trade in 1908, the wage distribution recorded by the Census of Manufactures in 1905, and numerous but scattered inquiries in narrow fields throw much light upon the subject, but the statistician is nevertheless compelled to make many rather sweeping assumptions on rather slender bases before he can

Digitized by Google

piece together a curve or table picturing income distribution among working class families.

Still further assumptions and generalizations are necessary before the curve for the working class can be coupled onto that for the well-to-do and rich, thus completing the money income distribution curve for the United States as a whole.

The problem of estimating total income for the people of the nation is relatively more simple, but cannot be accomplished with any high degree of accuracy by use of the existing data. Three separate lines of attack are available: First, the estimation of total consumption; second, the summating of net incomes of all individuals; and third, the calculation of the total annual product. Were these three quantities all computed with accuracy, it is evident that the last two would be nearly identical and that the first would be larger or smaller than the others according to whether the inventory of accumulated goods on hand was diminished or increased during the period considered.

To estimate total consumption, it is necessary to differentiate between goods used directly and goods used indirectly. It is necessary to evaluate the services of durable consumption goods and the direct services of persons. This includes such nice problems as estimating the value of poultry, vegetables, fruit, etc., consumed on farms; the aggregate rental value of residences and the amount spent for entertainment. Such estimates generally consist in little more than assuming as the correct figure the mid-point between possible limits of error.

Total book income for the nation can be estimated with a higher degree of accuracy than can total consumption. The former may be approximated by adding together estimated individual book incomes. Data concerning average wages in different occupations is quite abundant and the Census gives us a fair idea of the average number of persons following each calling. But even this process is not free from difficulties. Laboring families obtain much of their income from property owned, by renting rooms, by keeping boarders, etc. About the only study throwing enough light to be worth mentioning upon these supplementary income sources is the Eighteenth Annual Report of the Bureau of Labor, and this is 13 years

Digitized by Google

old. It is possible that conditions have changed noticeably during this interval. Add to this difficulty, the presumable underestimation of incomes in the federal and state tax lists of the wealthier classes and also the lack of information concerning the vast numbers of persons, such as farmers and small merchants who are not employees and are still too poor to pay an income tax, and the apparent simplicity of the problem becomes less evident.

The last recourse, then, in attempting to ascertain total national income is to turn to data on production. It is possible to find recorded values for most of the raw material and to trace the various value additions successively until the finished product finally emerges. But even this route is by no means all smooth sailing. We know the value added by manufacture and by transportation, but what about the value added by the great mercantile class? True, data are gradually accumulating concerning the fraction of the final selling price going to remunerate this function of adding time utility to goods but it is still complete in only a few lines. Likewise, the value of product contributed by personal servants, by the use of residences, churches, theatres, public buildings, etc., is nowhere accurately recorded. Yet, with all its defects, this is by far the easiest path to follow.

Would it not, however, be advisable for the Census Bureau to make direct inquiries concerning individual, family, and corporate income? Would it not be advisable to classify this income rather carefully as to source using a classification somewhat like the following:—

NET BOOK INCOME CLASSIFIED AS TO SOURCE.

1. Rent of land.
   a. Farm land.
   b. Mines.
   c. Urban land.
   d. Miscellaneous natural resources.
2. Rent of transition or capital goods.
   a. Business buildings and equipment.
   b. Miscellaneous.
3. Rent of direct or consumption goods, *e. g.*, residences.
4. Income from securities.
   a. Shares in partnerships or corporations.
   b. Notes, bonds, etc.

Digitized by Google

5. Income from wages or salaries.
6. Net entrepreneurial profits.
   a. Gain or loss in value of assets.
   b. Profits from operation of business.
7. Miscellaneous income.

It would be highly desirable, also, if we could learn something about the annual savings or deficit in the family finances, but it seems improbable that this difficult task could safely be added to the already heavy and intricate burdens of the census enumerator. Some will doubtless affirm that any inquiries concerning income are too inquisitorial to be entrusted to these functionaries with any hope of getting accurate results. The strength of this position is lessened by the fact that the class of people most reticent about their financial affairs have become accustomed to the inquiries connected with the income tax, and inquiries to which people are accustomed are much less likely to arouse their antagonism than are those of an unusual nature. There naturally would be numerous large errors but, as in the case of the study of wealth, these would probably tend to offset each other, since some would report too large and others too small a net income. The analysis of income made by the government for purposes of collecting the income tax would doubtless prove extremely valuable to those devising questions for a census schedule. While it might be necessary to make considerable modifications and alterations, the difficulties and mooted points have, at least, been clearly brought to light.

The income receivers would naturally be divided into individuals, partnerships, and corporations. The individuals ought probably to be reclassified on lines identical with those used for wealth. This plan would have the merit of facilitating comparisons of wealth and income.

IV.

Chronological Changes in Income. So much for the desired data on income which the Census might well attempt to gather in order to clarify the problems of distribution. After such an investigation, we could hope to answer more definitely questions not only concerning distribution among families but also

Digitized by Google

among the factors of production. These elaborate decennial studies are highly desirable in order, first, to give us a well-established base of operation and, later, to serve as guideposts to prevent us from wandering into paths of error. But, in nearly every field of economic activity, noteworthy fluctuations take place between the Census years. These fluctuations are largely cyclical in their nature and are of such great importance that the government can ill afford to ignore them to the degree which has been done in the past. To study these cyclical movements, we need not decennial, but rather monthly, or at most quarterly, reports. Evidently, this type of study must be carried on by enumerators who are always in the field. For such studies, complete data are not only forbidden by prohibitive expense but are absolutely unnecessary and useless. What is essential, is a collection of samples properly distributed both geographically and industrially.

The sampling for this purpose should be done on the theory of considering each citizen equally important. This means that samples should be collected from each state in proportion to the population; that the urban and rural districts should furnish samples in proportion to the inhabitants residing in each; and that the samples from each industry should be numerous in proportion to the number or workers which the industry employs. Were it not found feasible to apportion the samples in this fashion, the same results might be mathematically obtained by giving to each variety of samples the weights necessary to place all averages upon the basis of numbers of persons.

Stress is here laid upon the apportionment of samples or weights because the continuous statistical studies of the past have been given to overweighting certain classes. It is said that, in the last political campaign, the vision of certain politicians failed to reach beyond the Mississippi. Likewise, the visions of our governmental statisticians have too frequently failed to pass outside the city limits or even outside of certain factory districts. Continuous data concerning wages and hours of labor are confined largely to manufacturing industries and the unionized trades. Unemployment statistics are usually reported for trade unions only. The statistics for the

Digitized by Google

mines, the railways, and the rural districts are woefully inadequate and are almost incomparable with the results set forth by the federal Bureau of Labor Statistics. There is no intent here to reflect upon the competence of the people in charge of the respective fields of inquiry. They doubtless utilize their time and funds to as good advantage as circumstances permit. The present head of the United States Bureau of Labor Statistics deserves the highest commendation for the immense progress made toward putting the work upon a scientific basis, especially as regards the indices of wholesale prices. The criticism is on the lack of organization which permits different bureaus to gather statistics of similar nature and which allows purely technical positions such as those held by chiefs of statistical bureaus to be classified among the political spoils with the consequent likelihood of quadrennial changes in personnel.

But it is neither necessary nor desirable to discuss further at this time existing statistics. The fruitful thing is to ascertain what new data are really essential for measuring changes in income. Innumerable varieties are desirable which are of greatest importance. Those that seem to me fundamental are as follows:—

I. A General Weighted-Price Index to Measure Changes in the Value of Money with sub-index numbers as follows:

- A. Wholesale commodities.
- B. Stocks and bonds.
- C. Real estate, urban and rural.
- D. Rents, urban and rural.
- E. Labor per hour, classified by industries and occupations.
- F. Retail prices.

The data for this general index number should be weighted as nearly as possible in proportion to the volume of transactions in each field, in accordance with the principles laid down by Professor Irving Fisher.

II. The Hours of Labor per Week.

For this purpose all employees should be divided among the same industrial and occupational classes used for ascertaining the price of labor for the general price index described above. With information available concerning the price of labor per

Digitized by Google

hour and the hours worked per week, we are prepared to tell something concerning the fluctuations in the laborer's income.

III. A Consumer's Index of Prices, sub-divided according to the following classes of consumers:

A. Rich.
B. Upper middle class.
C. Lower middle class.
D. Poor.

The consumers' index should evidently have each article weighted in proportion to the amount thereof consumed by the given class. We have long possessed a satisfactory food index number. Beginnings are being made by the United States Bureau of Labor Statistics in recording the prices for dry goods and fuel. But dry goods are not the only kind of clothing, and price changes in so narrow a field may be far from typical. A large share of the ordinary consumer's budget fails to appear in any governmental price records. So far as the federal retail price indices are concerned, we do not live in houses, wear coats or shoes or hats, use electric lights, go to the theatres or movies, ride on the street cars or railway trains, read newspapers, or pay doctors' bills. In brief, there is no index number in existence which enables us to tell with any reasonable degree of accuracy whether either wages or income in purchasing power are going up or town. The constitution of such an average index number seems to me to be the most vital addition to our continuous statistics needed at the present time.

These figures should not only be gathered for the present but should, in so far as feasible, be carried back into the past. The obstacles to obtaining standard articles to represent each type of expenditures are great but they may be overemphasized. With patience and sufficient appropriations, it seems entirely practicable to present a consumers' index extending back for at least a couple of decades which would be accurate enough for all intents and purposes. After all, most people are not interested in minutiae and a statistician is, in no sense, duty bound to present only figures which are accurate to a high degree. The accountant must make his accounts balance to a cent. The statistician needs only to be sure that his figures

Digitized by Google

are accurate enough to answer *truthfully* those questions which they purport to answer at all.

IV. Employment, sub-divided under the same industries and occupations used for wages.

This fourth principal record is absolutely essential before we can judge changes in the income conditions of the working classes. It has hitherto been deemed impracticable to collect employment data from the workers themselves. This idea seems largely fallacious. In no other manner, can we ever hope to learn very much concerning lack of work among the unskilled and casual workers,—the classes who probably suffer most severely from this evil. No employer and no trade union secretary can tell us how many weeks these men were out of work during a given month or year.

To determine variations in the income of any part of the wage earning class we must then know four things:

1. The wage rate per hour.
2. The hours of labor per week.
3. The number of weeks employed per year.
4. The relative prices of the commodities purchased.

With these four records complete, we can picture, with a fair degree of accuracy, the changes in working-class income. The fact that there is much income from other sources is fully recognized, but the wage furnishes the only part in which there is marked variation.

It is, moreover, absolutely essential in any study of distribution and price changes to keep separate the above four ideas. To confuse the price of labor, which corresponds fairly closely to the hourly wage, with the daily, weekly, monthly, or annual earnings, all of which are essentially income concepts, is to sound the death knell of all accurate thinking in this field.

My knowledge of the methods used and difficulties involved in collecting this type of data is so meager as compared to that of many members of the American Statistical Association that I hesitate even to make a suggestion along this line, but I have often thought that it might be feasible to collect all of this information by aid of postmen. The city and rural mail-carriers penetrate into nearly every community of the country. They

Digitized by Google

know the people; they could easily ascertain the facts concerning prices, wages, hours of labor, and employment; they are men of intelligence; they all report to a central authority; why are they not the logical men to gather the needed data? Of course, they should have some slight extra compensation for performing this new task, but the expense involved should be comparatively trivial.

V.

Summary. The ideas which I have attempted to set forth here might be summed up as follows:

First, we need an enumeration of the most important problems to be solved.

Second, the facts deemed necessary to solve these problems should be accurately determined upon. It would be the legitimate function of a joint committee of the American Statistical and Economic Associations to outline both of the above fields in detail.

Third, the present duplication of work in these fields by different state and federal bureaus should be eliminated, all statistical work of this nature being placed under the direction of one central authority.

Fourth, this central bureau should so systematize the work as to give unity and continuity to the entire field. The principal averages of wages, prices, etc. should be presented monthly with a promptness and regularity approaching that which characterizes the appearance of the reports of the Weather Bureau or the index numbers published by financial magazines.

We are fond of boasting of American efficiency, and here seems to be an opportunity of demonstrating our claims in this line by putting into operation some such program as the one just outlined. If successfully executed, this would certainly set far to the front a new standard in the field of wealth and income statistics.

Digitized by Google

THE CLASSIFICATION OF MENTAL DISEASES IN NEW YORK STATE.*

By Horatio M. Pollock, Ph.D., *Statistician, New York State Hospital Commission.*

The preface of the first edition of the nomenclature of diseases, drawn up by a joint committee appointed by the Royal College of Physicians of London and published in 1869, declares that "for perfecting the statistical registration of diseases, with a view to the discovery of statistical truths concerning their history, nature and phenomena, the want of a generally recognized nomenclature of diseases has been felt as an indispensable condition." Further, that "among the great ends of such a uniform nomenclature must be reckoned that of fixing definitely for all places the things about which medical observation is exercised, and of forming a steady basis upon which medical experience may be safely built."

This committee was discussing the nomenclature of physical diseases, but its observations apply equally well to mental diseases. The lack of a generally accepted classification of mental diseases makes it impossible at the present time to collect comparative data concerning the various forms of mental diseases in different states and countries. In taking the census of the insane in institutions in 1910, the federal Census Bureau found it impracticable to collect data concerning any specific forms of insanity, except the alcoholic psychoses and general paralysis, and the data collected concerning these are in part unreliable.

In New York State, since 1889, four different systems of classification of mental diseases have been used. The first, a modification of the English classification, was used in a perfunctory way, but with no pretense of scientific accuracy, from 1889 to 1902. Then a simplified form of the same classification was introduced, but this was no better applied than the first. In 1905 the new psychiatry was rung in and the old

* Read at the quarterly meeting of the American Statistical Association, New York City, December 14, 1916.

Digitized by Google

classification was rung out. The process, however, was not joyous. The physicians in the service had had a comfortable time with the old classification. To apply it required but little study of cases, and very few persons thought of comparing the diagnoses of one hospital with those of another. The new classification required study and close observation, and the work done in the hospitals was likely to be checked.

The attitude of these physicians was well shown in the remarks made by Dr. A. E. Macdonald, then superintendent of the Manhattan State Hospital, in his presidential address before the American Medico-Psychological Association in May, 1904. Speaking of matters of importance to the association, he said: "One of these is the perennial question of the classification of insanity, which, often as it has been agitated and pondered, has yet failed of satisfactory adjustment. It is a matter of great regret that some at least working agreement can not be reached, faulty even though it should be, which will enable the alienists of one country to understand the statistics of others, and to apply them by way of comparison to his own. If such a standard *is* to be reached, it would appear to me that it must be through mutual concessions and agreements of practical men, such as compose our own and kindred associations, for I opine that present conditions result from less possible coherent elements: authors and clinicians for example, who have pride in their own classifications, unstable though they be, and are incapable of recognizing possible value in others." He then referred to the history of New York State, and added, referring to the new classification, that "we are threatened with still another revolution, and that in the direction of intricate association of newly discovered or invented forms which promise little in the way of adhesiveness or permanency."

Judging the matter in the light of subsequent history, Doctor Macdonald was a better statistician than psychiatrist. He realized the need of a systematic classification of diseases that would have permanency of character, but evidently had little conception of how such a system was to be devised. In contrast with the standpatters, it is interesting to note the attitude of Dr. Adolf Meyer, one of the so-called "authors and clini-

Digitized by Google

cians," who, as director of the Psychiatric Institute, formulated the new classification in 1905. Speaking of the system of statistics, which the president of the American Medico-Psychological Association thought was fairly satisfactory, he said: "The present statistics of forms of mental diseases are of very little use. However comparable they may appear, they have merely sham value, as could easily be shown in actual cases, such as in many of manic-depressive insanity, which have figured under no less than the following headings: Acute mania, acute melancholia, chronic mania, chronic melancholia, secondary dementia, paranoia, mania recurrent, melancholia recurrent, and circular insanity, where the same kind of attacks recurred, or on the arbitrary differences which exist in the use of the term paranoic and acute and simple melancholia, etc. between the various hospitals. The hospital physicians do not consider themselves responsible because the system is forced on them and they submit because they have nothing to offer which would be accepted by all."

In discussing the difficulties of classification, he stated that an absolutely clean cut classification is not to be expected, unless the number of groups be made unwieldy, and further explained that improvement in classification would come with the increase of clearly definable facts. Although Doctor Meyer's classification was adopted by the commission in 1905, it was not till 1908 that it was used with any degree of success throughout our state hospital service. During the interim of three years the well recognized organic types were diagnosed and counted, but the remainder were reported under the heading "all other psychoses."

In 1908 the Commission in Lunacy appointed a statistician, and made provision for the reporting by the hospitals of definite statistical data relative to each patient admitted or discharged. Such data included among other things the diagnosis of the mental disease with respect to group and type. Distinction was made between first admissions and readmissions. The per cent. distribution of psychoses in the various hospitals, compiled from the statistical cards submitted for both classes of admissions, has been compared from year to year, and discussions concerning variations have been held at

Digitized by Google

several quarterly conferences. These reports and discussions have had the effect of making the physicians in the several hospitals cautious in diagnosis, and have compelled close study of the history and mental and physical characteristics of the patients admitted. The excellent results now being obtained, however, would not be possible had it not been for the courses of instruction in psychiatry given to the physicians in the hospital service by Doctors Meyer and Hoch of the Psychiatric Institute. These courses have given physicians in the service clearer conceptions of the new psychiatry, and to some extent at least the same point of view with reference to classification.

In 1915 it was found that the classification in use needed some revision to bring it up to date, and after a conference with the committee on statistics, it was modified by Doctors Hoch and Kirby. At the present time the classification conforms fairly well with the best modern psychiatry.

Doctor Meyer, in making the orginal classification, started with three large groups, viz.:

A. The diseases which depend on a fairly tangible interference with the brain or its nutrition, in the form of some definite brain disease, or some intoxication, or some autotoxic or infective or exhaustive disorder.

B. Disorders which provisionally, and owing to their more or less striking symptomatology, are most interesting from the point of view of *symptomatic equivalents, and course and outcome;* simple excitements and depressions, dementia praecox, paranoic developments, and manic-depressive states.

C. Disorders in which we see the marks of some of the recognized neuroses, hysteria, epilepsy, neurasthenia, psychasthenia, and peculiar make-up; and inbecility and idiocy.

The first large group A, was subdivided into smaller groups; one comprising the definite brain diseases, one the intoxications, and a third, infective and autotoxic disorders.

In the first sub-group the psychoses are closely related to an organic, or at least definite disease, or disturbances of the brain and nervous system. Clinical groups under this subdivision in the classification now in use, are the traumatic psychoses, psychoses with cerebral arteriosclerosis, dementia paralytica,

Digitized by Google

or general paralysis, psychoses with cerebral syphilis, psychoses with Huntington's chorea, psychoses with brain tumor, and psychoses with other brain or nervous diseases.

The second sub-group, the intoxications, includes the alcoholic, drug and other toxic psychoses. The general group of infective and autotoxic disorders comprises infective-exhaustive psychoses, allied to infective-exhaustive psychoses, and autotoxic psychoses.

The second large group of Doctor Meyer's classification, B, which is generally spoken of as the functional psychoses, includes in our present classification seven clinical groups, viz.: manic-depressive, allied to manic-depressive, involution melancholia, symptomatic depressions, dementia praecox, allied to dementia praecox, and paranoic conditions. Owing to the fact that these clinical groups are not fully understood, classification of borderline cases is exceedingly difficult. Certain cases that do not present the typical symptom-pictures of the more distinct psychoses are placed in the allied groups.

In the third of Doctor Meyer's large groups, C, are now placed the epileptic psychoses, psychoneuroses, psychoses with constitutional inferiority, and psychoses with mental deficiency.

Although Doctor Meyer was a pioneer in the introduction of the new psychiatry in this country, he was successful in introducing a system that has proved its worth during the past eight years in the New York State hospitals. When the advancement of science brings about a marked change in the conception of some forms of mental diseases, our classification will be modified accordingly; in the meantime we have the satisfaction of living up to the light we have.

In the treatment and study of mental diseases, New York State has always been a pioneer. It was first among the states to establish a comprehensive and enforced system of state care of the insane; first to provide separate institutions for the criminal insane; first to establish a scientific institute for research and instruction in psychiatry; first to conduct training schools for nurses in its hospitals for the insane, and first to

Digitized by Google

establish a bureau of statistics for the systematic study of data relating to insanity.

Other states have followed New York in many things, and it is hoped that its system of collecting scientific data of mental diseases will also be found worthy of general adoption.

I append a copy of the classification now in use, with some explanatory notes prepared by Drs. Meyer, Hoch, and Kirby.

CLASSIFICATION OF MENTAL DISEASES ADOPTED BY THE NEW YORK STATE HOSPITAL COMMISSION SEPTEMBER 22, 1915, TO TAKE EFFECT OCTOBER 1, 1915.

1. *Traumatic psychoses.*

"We should not speak of traumatic psychoses where the trauma is merely a contributory factor in precipitating another definite mental disorder such as general paralysis, manic-depressive insanity, katatonic or paranoic deterioration without specific traumatic stigmata, or where a psychosis follows an injury not involving the brain, through shock, etc." (Meyer.)

The following are the most common clinical forms:

(a) Post-traumatic delirium.
(b) Post-traumatic constitution (irritability, sensitiveness to alcohol, paranoid trends, hysteroid or epileptoid attacks).
(c) Traumatic defect conditions (aphasia, deterioration with epilepsy, mental enfeeblement, etc.).

2. *Senile psychoses.*

The clinical types most frequently met with are the following:

(a) Simple deterioration (of memory, intellectual capacity and interests).
(b) Presbyophrenic type (similar to Korsakow mental complex).
(c) Delirious and confused states.
(d) Depressed and agitated states in addition to the deterioration.
(e) Paranoid states in addition to the deterioration.

3. *Psychoses with cerebral arteriosclerosis.*
4. *Dementia paralytica.*
   (a) Cerebral form (exaggerated knee-jerks).
   (b) Tabetic form (diminished or absent knee-jerks).
5. *Psychoses with cerebral syphilis.*
6. *Psychoses with Huntington's chorea.*
7. *Psychoses with brain tumor.*
8. *Psychoses with other brain or nervous diseases.*

This division provides a place for grouping a variety of more rare mental disorders associated with organic diseases of the nervous system. On the card the special disease must be mentioned after the group name, and this should be followed by a few descriptive terms, characterizing the accompanying mental disturbance.

(a) Cerebral embolus.

Digitized by Google

(b) Brain abscess.
(c) Tubercular meningitis.
(d) Central neuritis.
(e) Multiple sclerosis.
(f) Tabes dorsalis.
(g) Acute chorea.
(h) Other conditions to be specified.

9. *Alcoholic psychoses.*

Since on the basis of chronic alcoholism a number of fairly characteristic mental disturbances develop, the type of alcoholic psychosis should be indicated in each case. Among these types only "pathological intoxication" seems to require special explanation. By the term "pathological intoxication" is meant a condition which is the *immediate* result of taking a larger or smaller amount of alcohol. The symptom-picture is characterized by special features not seen in ordinary drunkenness.

(a) Pathological intoxication.
(b) Alcoholic deterioration.
(c) Delirium tremens.
(d) Korsakow's disease (polyneuritic psychosis).
(e) Acute hallucinosis.
(f) Chronic hallucinosis.
(g) Acute paranoid form.
(h) Chronic paranoid form.
(i) Other less characteristic symptomatic types should be specified.

10. *Drug and other toxic psychoses.*

(a) Morphinism, cocainism, bromism, chloralism, etc., or combined.
(b) Lead intoxication (several forms are described).
(c) Gas poisoning (delirium or confusion).
(d) Food toxicoses (tea, coffee, ergot, etc.) and pellagra.

11. *Infective-exhaustive psychoses.*

(a) Febrile delirium.
(b) Infection delirium.
(c) Exhaustion delirium (acute confusion, hallucinatory delirium, "amentia").
(d) Delirium with heart disease.
(e) Other conditions to be specified.

12. *Allied to infective-exhaustive psychoses.*

Cases which clinically resemble infective-exhaustive states but in which no infective-exhaustive etiology can be found.

13. *Autotoxic psychoses.*

(a) Thyreogenous disorders.
(b) Uremic disorders.
(c) Diabetic disorders.
(d) Other conditions to be specified.

14. *Manic-depressive psychoses.*

(a) Manic attack.
(b) Depressive attack.

Digitized by Google

(c) Stupor.
(d) Mixed attack.
(e) Circular attack.

15. *Allied to manic-depressive psychoses.*

16. *Involution melancholia.*

In spite of the fact that many of the depressions which have thus been designated are related to manic-depressive insanity, and that therefore a differential diagnosis may seem rather unnecessary, it is nevertheless advisable to group here all the depressions past middle life which show plain anxiety and agitation, and regard them as *special* form of the emotional reactions.

17. *Symptomatic depressions.*

Depressions closely associated with some physical ailment which more or less naturally predisposes to despondency (heart disease, gastric disorders, malignant growth, etc.). Fluctuations occur in the depression depending largely on the severity of the physical symptoms or degree of discomfort or pain which the patient suffers.

18. *Dementia praecox.*
(a) Paranoid form.
(b) Katatonic form.
(c) Hebephrenic form.
(d) Simple form.

19. *Allied to dementia praecox.*

20. *Paranoic conditions and paranoias.*

Delusional conditions in which the train of thought is clear and the contact with the environment is well preserved, though episodes of greater disturbance may occur.

21. *Epileptic psychoses.*

The epileptic disorders may be grouped according to the most prominent clinical manifestations, as follows:
(a) Deterioration.
(b) Clouded states.
(c) Other conditions.

22. *Psychoneuroses.*

These essentially substitutive reaction types may be outlined as follows:
(a) Hysterical type.

Submersion of a disturbing experience or conflict and conversion into hysterical manifestations, as a rule in the nature of somatic symptoms.

The psychoses most frequently characterized by psychogenic attacks, with more or less clouding of consciousness with subsequent amnesia.

(b) Psychasthenic type.

Conditions with phobias, obsessions, impulses, doubts, often associated with tension.

(c) Neurasthenic type.

Great mental and motor fatiguability accompanied by irritability, pains, hyperasthesias. The neurasthenic type can not always be satisfactorily differentiated from the psychasthenic.

Digitized by Google

23. *Constitutional inferiority with psychoses.*

By constitutional inferiority we mean abnormalities of mental make-up in which defects exist in the emotional and volitional sphere, but in which there is no obvious intellectual deficiency. Therefore we might in many cases of the well known constitutional psychoses speak of the individuals as constitutionally inferior (*e.g.* in dementia praecox or in manic-depressive insanity) but for obvious reasons it is better to limit the term "constitutional inferiority" to types of abnormal mental make-up which differ from these and in which the psychotic reactions also differ from the psychoses already specified in earlier groups.

24. *Mental deficiency with psychoses.*

This includes intellectual deficiency of various degrees. Traits of constitutional inferiority may be added.

25. *Unclassified.*

In this group should be included cases in which a diagnosis can not be made owing to lack of adequate data; or cases in which the clinical picture is not understood and the etiology is not known.

26. *Not insane.*

(a) Epilepsy.
(b) Alcoholism.
(c) Drug addiction.
(d) Constitutional inferiority.
(e) Mental deficiency.
(f) Dotage.
(g) Others to be specified.

Digitized by Google

THE INCREASING MORTALITY AFTER AGE FORTY-FIVE—SOME CAUSES AND EXPLANATIONS.*

By Louis I. Dublin, Ph.D., *Statistician, Metropolitan Life Insurance Company, New York.*

It is important that I state the problem of this paper clearly at the very outset. American statisticians have been concerned for some years with the fact that the death rates after age 45 are higher than they were, say, 10 or 15 years ago. Further back than 1900 it is impossible to go because of the lack of accurate data. The figures for the registration states in 1900 compared with those for the same states in 1914† show substantial decreases for every age period and for both sexes up to age 45. After age 45 there is a slight increase in the death rate for males. It is less than 4 per cent. in the age period 45–54; about 10 per cent. in the age period 55–64, and about 4 per cent. in the period 65–74. The age period 75 and over, shows little of importance. For females, the increase does not begin to appear until the age period 65–74. It is then about 1 per cent. higher in 1914 than in 1900. The reader will note that these increases are not very marked. They would call for very little attention were it not for the fact that this 15 year period has been one of extensive development in our public health service and of increasing prosperity in our country. It is perhaps disconcerting that there should have been an increase at all since the mortality rates of the more important European countries have shown no such increases at the corresponding ages.

The problem becomes much more interesting and acute as we leave the total mortality and concern ourselves with that from some of the causes of death which play an important part after age 45. This is, indeed, the phase of the subject that has received the greatest publicity and attention. We find, for example, in the registration area, in the registration states

* Read before the Annual Meeting of the American Statistical Association, Columbus, Ohio, December 28, 1916.

† The death rates for 1914 (registration states as constituted in 1900) were computed on the basis of the estimated population for 1914 distributed by age and sex in accordance with the 1910 census.

Digitized by Google

and in selected states having the longest period of good registration, that such causes of death as cancer, diabetes, cerebral hemorrhage and apoplexy, organic diseases of the heart, diseases of the arteries, cirrhosis of the liver, and Bright's disease, all show higher death rates in 1914 than in 1900. These causes of death are characteristic of the older age periods of life and in 1914 together accounted for 58.7 per cent. of the total mortality after 45 in the registration states as constituted in 1900. Cirrhosis of the liver shows the least increase in the total death rate, about 10 per cent. in these 15 years. Diseases of the arteries show the highest increase, nearly 454 per cent. Diabetes shows an increase of nearly 76 per cent. and organic diseases of the heart close to 45 per cent. in the same period. This then is our problem: to attempt to identify the factors involved in these higher death rates and to offer an explanation.

Some interesting explanations of the increasing mortality at the higher age groups have already appeared. The increasing industrialism of the country is supposed to subject the individual workman to greater wear and tear, which shows itself in higher death rates after age 45. The influence of the growth of our cities at the expense of the rural districts and the increasing number of persons engaged under the factory system—all of these points have been emphasized. Changing personal habits of life have also been charged with an important part in the phenomenon. It has been suggested that the supposed harder pace of business and social life today, the possibly diminished opportunities to get out in the open, to exercise, and to live moderately, have had an unfavorable effect upon those over 45. This has been the explanation offered by those identified with the Life Extension movement and by the advocates of periodic physical examinations generally. They have found in this thesis a powerful motive for their very useful propaganda. Their statements have in common the basic assumption that there has been a real and significant deterioration in our national vitality at the ages under consideration.

It is my purpose today to urge great caution in our acceptance of this conclusion. As statisticians we should first be sure that the simpler logical requirements of the problem have

Digitized by Google

been met before we attempt explanations at all; for with the field cleared up there may be no real phenomenon to explain. This is in effect what I propose to show.

In making comparisons we must make certain of the comparability of our data. In terms of the immediate problem, this means that we must be certain that the mortality rates of the earlier period, say 1900, are for a group of persons similar in essential respects to those of the later period. As a general thing, figures for the European countries meet this requirement. If the reports of the Registrar General of England and Wales show that the death rate per 1,000 in the age period 45–54 was 16.8 in 1880 and in 1910 only 12.7, we must conclude that there has been a decrease of 24 per cent. in the mortality at that age period and that the conditions of life have improved to that extent. But we cannot draw conclusions either one way or the other from similar American figures and for this reason: the character of our population of the earlier period has not remained unchanged. The years which we are considering were years of great change in the race constitution of American states and cities. They were years in which occurred a great influx of newer race stocks, an immigration which is making up a large and ever-increasing part of the population of our registration area. It is indispensable to the argument that the effect of this change in our race stock shall be determined as a part of the phenomenon we are considering. Does it make any difference that we have added to our population millions of Russians and Italians and Poles, whom we did not have at the earlier date? If not, then clearly the increasing mortality means something; it might even indicate physical deterioration after 45 as some writers would have us believe. If it can be shown, however, that these foreign races are contributing to the older ages of our population and that at these ages they have consistently higher mortality rates than our native stock, then the increase in the total mortality indicates nothing as an index of deterioration in the vitality of our native population.

The question resolves itself thus: Are *Americans* today experiencing a shorter expectation of life after 45 than did

Digitized by Google

Americans of a previous generation? I mean by Americans the native born of native parentage. Statistically, the problem has no meaning or value when stated in terms of the total population. The problem must be analyzed in terms of homogeneous groups of the population and comparisons must be limited within these groups.

For some time I have interested myself in just this phase of the problem. In a recent study of the death rates in the race stocks in New York State, 1910,* I showed how the important races compared with one another in respect to their mortality. Comparisons were made between (1) the native born of native parentage, (2) the native born of foreign or mixed parentage, and (3) the foreign born. Statistics are given in full in that report, but for this occasion, I shall summarize only those results which bear upon the problem of middle and old age mortality. We found consistently that after age 45 the group of native born of native parentage had the lowest mortality. This was true for both sexes. The superiority of the native stock over the foreign born is very marked indeed. The foreign born males had a death rate of 28.0 per thousand in the age period 45–64 as against a rate of 18.8 for the native born of native parentage, which is a difference of about 49 per cent. in favor of the latter group. The foreign born females in the age period 45–64 showed a rate of 23.4 as against a rate of 14.3 for the native born, a difference of nearly 64 per cent. in favor of the native born. The foreign born and the native born of foreign or mixed parentage, were remarkably alike as to the mortality of the males. This was not so for the females, however, the foreign born showing the highest rates.

On further analysis of the foreign born mortality we were able to distinguish the characteristics of the following nationalities: (1) Russians (for the most part Jews, in New York State), (2) Italians, (3) Germans, (4) Irish, (5) Austro-Hungarians, and (6) English, Scotch, and Welsh. Whatever be the foreign stock, virtually the same phenomenon of higher mortality appears in comparison with the native born of native parentage, the differences between the several nationality classes being one of degree only. Even the Russians

* "Factors in American Mortality—A Study of Death Rates in the Race Stocks of New York State, 1910." American Economic Review, Vol. VI., No. 3, September, 1916.

Digitized by Google

(Jews mostly) who are notable for their very low death rates during the early working years of life show consistently higher death rates after age 45 than do the native born of native parentage. When we turn to the Germans and the Irish the differences are remarkable. Thus, the Irish at the period 45–64 showed a rate of 46.3 per 1,000 for men as against 18.8 for the native born of native parentage, and for women, a rate of 40.7 as against 14.3 for the native born women of native parentage. The following table gives the figures for each of the foreign race stocks in comparison with the native born of native parentage for the age periods after 45 and for each of the sexes:

TABLE I.

DEATH RATES PER 1,000 IN PRINCIPAL NATIVITY CLASSES OF NEW YORK STATE POPULATION, 1910, CLASSIFIED BY AGE PERIOD AND BY SEX.

Age Period and Sex.	Native Born of Native Parentage.	Native Born of Foreign or Mixed Parentage.	Foreign Born.	Persons Born in					
				Russia.	Italy.	Germany.	Ireland.	England, Scotland, Wales.	Austria-Hungary.
45–64									
Males . . .	18.8	28 2	28.0	20.1	19.3	27.7	46.3	24.6	21.0
Females .	14.3	20.0	23.4	16.0	17.9	18.4	40.7	21.0	18.2
65–84									
Males . . .	77.3	89.9	90.4	78.4	64.6	90.4	101.6	86.6	77.5
Females .	68.2	73.9	87.7	66.8	63.8	83.1	107.4	79.9	63.9

This evidence cannot be discounted when we consider the problem of the increasing mortality after age 45. Its significance is evident when we note the changes that have occurred in the racial composition of the population of New York State during the 10 year period 1900 to 1910. In the former year, 39.2 per cent. of the entire population was native born of native parentage; in the latter year this proportion had fallen to 35.4 per cent. On the other hand, the foreign born white population increased from 26.0 per cent. to 29.9 per cent. The native born of foreign or mixed parentage changed slightly from 33.2 per cent. to 33.0 per cent. The following table shows side by side the proportionate distributions in 1900 and in 1910 of the three main groups of the population in New York State in the several age periods after 45:

Digitized by Google

TABLE II.

PERCENTAGE OF GENERAL NATIVITY GROUPS IN THE WHITE POPULATION OF NEW YORK STATE AT CERTAIN AGE PERIODS, 1900 AND 1910.

Age Period and Population Group.	1910.	1900.	Difference 1910–1900.
45–54 Years			
Total white population	100.0	100.0	....
Native born, native parentage	32.0	39.3	−7.3
Native born, foreign or mixed parentage	26.6	18.6	+8.0
Foreign born	41.4	42.1	−.7
55–64 Years			
Total white population	100.0	100.0	....
Native born, native parentage	39.8	44.4	−4.6
Native born, foreign or mixed parentage	18.0	10.3	+7.7
Foreign born	42.2	45.3	−3.1
65–74 Years			
Total white population	100.0	100.0	....
Native born, native parentage	45.9	49.3	−3.4
Native born, foreign or mixed parentage	10.5	5.8	+4.7
Foreign born	43.6	44.9	−1.3
75–84 Years			
Total white population	100.0	100.0	....
Native born, native parentage	51.2	56.6	−5.4
Native born, foreign or mixed parentage	6.3	4.8	+1.5
Foreign born	42.4	38.7	+3.7

The above table shows that the main effect of the change in the composition of our population has been to decrease the proportion of the native born of native parentage, that is, the group with the most favorable mortality after age 45. In 1900 this group comprised 39.3 per cent. of the New York State white population aged 45–54, but by 1910 this proportion had dropped to 32.0 per cent. Similarly in the age period 55–64, the proportion of the same group dropped from 44.4 per cent. to 39.8 per cent. The decrease in the proportion of the foreign born is more than balanced by the increase in the native born of foreign or mixed parentage who, as shown in Table I, are much more similar to the foreign born in their mortality characteristics than to the native born of native parentage.

An increase in the proportion of those having higher death rates and a corresponding decrease in the proportion of those having lower death rates must have the effect of increasing the total death rate. From this conclusion there can be no escape. The problem is, indeed, analagous to that of age standardization which is a necessary and accepted procedure in every day vital statistics work. No qualified statistician would com-

Digitized by Google

pare the crude death rates of two communities without taking into account the character of their age constitution at least. The Census Mortality Reports, for example, specifically give warning of the dangers of not making such corrections. It might very well be that one population had a larger proportion of infants and old people than the other with which it was being compared. I now insist that it is equally fallacious to compare the populations of two periods even for the same state when we know that there has been a change in the race composition and especially when we also know that this change affects the total death rate. In this particular instance, the change in the race composition in New York State has certainly increased the mortality after age 45.

So much for the death rates for all causes combined. The argument runs along very similar lines with reference to the supposed increase in the rates of the more important of the individual causes of death referred to above. The following table shows how much higher the death rates in New York State for two of these causes are for the foreign born than for the native born of the same ages. There are some exceptions, to be sure, but they are not many nor are they very important.

TABLE III.

DEATH RATES PER 100,000 FROM BRIGHT'S DISEASE AND ORGANIC DISEASES OF THE HEART FOR THE PRINCIPAL NATIVITY CLASSES OF NEW YORK STATE POPULATION, 1910.

Cause of Death, Sex and Age Period.	Persons Born in						
	United States.	Russia.	Italy.	Germany.	Ireland.	Austria-Hungary.	England, Scotland, Wales.
Bright's disease							
Males							
45–64	267.2	238.6	172.9	332.4	662.5	298.6	288.3
65–84	952.2	1322.4	807.9	1306.7	1299.1	1628.9	861.4
Females							
45–64	208.6	276.7	211.5	245.2	636.1	241.0	241.9
65–84	680.9	1273.6	549.8	1122.3	1388.6	852.6	826.0
Organic diseases of the heart							
Males							
45–64	316.3	254.2	250.8	328.9	580.2	244.3	272.1
65–84	1693.7	1254.0	1303.1	1905.5	2088.5	1062.3	1903.0
Females							
45–64	247.1	239.4	373.6	288.5	656.4	221.3	295.6
65–84	1369.3	1485.8	1649.3	1766.5	2312.1	1136.8	1791.4

Digitized by Google

It is unfortunate that I cannot give you for comparison the figures for the native born of native parentage for these causes. If we had reliable figures for this group the differences shown would stand out in even greater relief. This is clearly indicated in certain data now in our possession which are not yet ready for presentation. The rates for Bright's disease in the age period 45–64 are lowest for males born in the United States except for those born in Russia and in Italy. Males born in Ireland give a rate more than twice that for natives. Among females, ages 45–64, the rate for this disease is without exception lowest for the native born. In the age period 65–84 an occasional exception appears for both males and females; but, taken by and large, Bright's disease is more prevalent at these advanced ages among the foreign born stock than among natives. The figures for organic diseases of the heart, and, as reference to my previous study, "Factors in American Mortality," will show, the figures for other of the degenerative diseases point to a similar conclusion.

The apparent increase in the mortality from these causes in the population since 1900 must, therefore, be considered in the light of the changing character of our race constitution. The facts brought out in the discussion of the general death rates above would seem to have an even greater application in relation to these special causes of death.

I desire finally to consider the degenerative disease problem from another angle. It may throw additional light on the supposed increase in their death rates. The diseases or rather the causes of death are those previously mentioned, but I shall repeat them for the sake of clearness: (1) Cancer, (2) diabetes, (3) cerebral hemorrhage and apoplexy, (4) organic diseases of the heart, (5) diseases of the arteries, (6) cirrhosis of the liver, and (7) Bright's disease. The supposition has been that the true incidence of these diseases has increased because their rates have gone up more or less steadily since 1900. I have shown above that the changing race factor played an important part in this rise, but in addition we must remember that the published rates are more or less subject to the influence of improving technical procedure in statistical offices. This is

Digitized by Google

not a minor matter. It is especially true of this group of diseases that the methods of classification in the Bureau of the Census and in other registration and compiling offices have been subject to progressive change. This has been the result of decisions of the International Commission and of the ever increasing effort to obtain more specific and complete statements of the primary cause of death. Moreover, many of the diseases under consideration are subject to errors of diagnosis. They often occur together and in combination with other conditions. The routine of stating them on death certificates as causes has not even now become crystallized in the minds of physicians. The last 15 years, moreover, has been a period of increasing interest in these diseases with the result that they are mentioned more often in cases where formerly no reference to them would have been made. All of these factors combined serve to make the group a very unsatisfactory one to discuss on the basis of the bare figures. There is a great deal to explain before comparisons can be made between 1900 and 1914. I propose in the following to take up a number of these diseases and to show how seriously the rates for them have been modified solely as the result of changes in medical and statistical nomenclature and practice.

I. CANCER.

The published rates for cancer have been markedly increased by more definite statements of cause on death certificates. Deaths previously reported as due to "tumor" and assigned to the non-malignant tumor class are now more generally described as cancers of some definite organ. This, moreover, is confirmed by the actual experience of those who have handled large numbers of death certificates and who in their correspondence with physicians have observed how readily a statement of "tumor" can be converted into a statement of cancer as a result of inquiry. We are also charging today many deaths to cancer which, in 1900, on the basis of the descriptions then given by physicians, would have been classified as septicemia, hemorrhage, pulmonary congestion, peritonitis, acute nephritis, and other conditions. This is also true of such titles as "disease" of the brain, "disease" of the

Digitized by Google

larynx, the mouth, the pharynx, the esophagus, the stomach and various other organs and members of the body usually subject to cancerous growths. Moreover, there is today much more certainty than before in the diagnosis of cancer as the result of the increase in the number of operations and laboratory examinations. The result of these changes in statistical and medical practice has been an increase in the reported death rate from cancer.

II. CEREBRAL HEMORRHAGE AND APOPLEXY.

Much of the apparent increase in cerebral hemorrhage and apoplexy is accounted for by the decrease in reports of "paralysis." In 1900, the rate for the latter disease was 25.9 per 100,000; in 1914 it had fallen to 9.7. Physicians now less frequently report "paralysis" but instead give the cerebral hemorrhage which caused the paralysis.* Furthermore, it is now the practice of vital statistical offices to return certificates where "paralysis" and "hemiplegia" are reported as causes of death and many such cases come back to the registration offices changed to "cerebral hemorrhage." Thus, in the returns of the Statistical Bureau of the Metropolitan Life Insurance Company, 61.1 per cent. of the queries on "paralysis" resulted in changes to "cerebral hemorrhage and apoplexy." These facts together with the known changes in race composition referred to earlier lead us to doubt seriously whether there has been any real increase in the incidence of this disease. All things considered, therefore, I doubt if there is any basis for the conclusion that there has been an actual increase in the mortality from this cause of death.

III. ORGANIC DISEASES OF THE HEART.

The marked changes that have occurred in the statistical treatment of this cause of death are best indicated by a comment appearing in the report of the Bureau of the Census on Mortality Statistics for the year 1910, page 54. In discussing the increase in the published rate from this disease the report says: "A part of this increase, however, was only apparent, being due to the fact that prior to the second revision of the

* In 1900, of the total number of deaths reported from both "paralysis" and cerebral hemorrhage and apoplexy, 27.6 per cent. were ascribed to the former; in 1914 only 11 per cent. were so returned.

Digitized by Google

International List, all deaths reported as due to 'endocarditis' (unqualified) and 'myocarditis' (unqualified) were compiled under 'acute endocarditis' (International Title No. 78), whereas in the compilation of 1910 the classification of the second revision has been used and such deaths of persons 60 years of age or over have been classified under 'organic diseases of the heart.'" This change in the classification of endocarditis and myocarditis has been in force since 1910.* Change in accepted procedure in editing jointly returned causes of death has also played a part in this seeming increase in the mortality from organic heart diseases. The Index of Joint Causes of Death published by the Bureau of the Census in 1913 shows that specifically reported causes like "mitral insufficiency," "valvular disease of the heart," etc., are now made preferred causes over certain diseases such as intestinal obstruction, paralysis agitans, tubo-ovarian abscess, Ludwig's angina, renal calculus, elephantiasis, and others. These, on the other hand, when reported in combination with less definite terms like "heart disease" are made preferred causes over the latter. Each year fewer physicians are returning "heart disease" and more of them are returning such definite reports as "mitral insufficiency." Finally, many deaths which formerly were returned as due to "dropsy" and to other conditions which were classified under the general heading "ill-defined organic disease," have lately been definitely ascribed to organic diseases of the heart. "Ill-defined organic diseases" have dropped from 7.2 per 100,000 in 1900 to 0.8 in 1914. These causes are especially prominent in the older ages. The consequence is that, more and more, deaths are being charged to organic heart diseases which formerly would have been tabulated elsewhere.

IV. DISEASES OF THE ARTERIES.

The immense increase in the published rate for diseases of the arteries is obviously not in accord with the facts. We are concerned simply with artifact—a change in the practice of physicians in reporting the causes of death in old age. It is

* In fact, in each report on mortality since and including that for 1910 the Bureau of the Census has placed on title No. 79 (Organic diseases of the heart) in its table giving rates for diseases for several years an asterisk to indicate that there have been changes "markedly affecting the comparability of the figures."

Digitized by Google

the experience of persons who have been engaged for years in handling death certificates that deaths which are now charged to arteriosclerosis would often have been reported in former years as due to "old age." In 1900, the death rate for "old age" was 50.4 per 100,000 and in 1914 it was only 19.8, that is, less than two fifths of the earlier rate. The interest of the medical profession has in recent years been directed toward arteriosclerosis and to allied diseases with the result that these diseases are more often reported where previously little attention would have been given to them. Hardened arteries may be assumed to be a normal condition in old age rather than a morbid one and it is partly because of the newer interest in the etiology of the condition that more deaths are being reported as due to them.

The progress that has been made in defining statements of cause of death is shown to some extent by the marked reduction in the death rates of all causes which are vague and indefinite in their nature. We have already referred to the decline in the rate for the ill-defined diseases. Deaths which were previously referred to this unsatisfactory group of causes are now very commonly classified at the older ages under more specific and definite titles, especially under one or another of the so-called "degenerative diseases."

SUMMARY AND CONCLUSION.

On the main subject at issue I am constrained to conclude that there is no evidence to support the contention that the true mortality rate of the country has increased after age 45. The problem will be decided definitely only when a new set of life tables, similar to those issued by the Census Bureau for the year 1910, appear at later decennial periods. No such data are now at hand. Enough facts are in evidence, however, to warrant the conclusion that the changing constitution of our race stock is responsible for the slight increase in mortality that has occurred. It is perfectly true that the marked decreases in mortality which are characteristic of the earlier ages of life do not continue after age 45. Nor could they be expected to under any circumstances. Immortality is not one of our characteristics. Persons who do not die from the in-

Digitized by Google

fections and accidents at the earlier ages will necessarily come to death at the more advanced ages. This is no cause for apprehension from the health or any other community standpoint. We are not concerned with a social misfortune when the mortality after age 75 increases and apparently the figures indicate that for the present at least no increases have occurred up to that advanced age.

As to the alleged increase in the death rate from the so-called degenerative diseases, the causes of death most prominent in middle and old age, these are due to two factors; first, to the changes in our race composition, and second, and perhaps equally important, to the changes in International and local statistical practice with reference to their assignment on death certificates. Together, these facts throw grave doubt on the assumption that any of these diseases has increased at the ages in which we are economically and socially interested. It is high time, therefore, that the ghost of these degenerative diseases was laid and the program for life conservation at middle life and at the older ages was established upon sound scientific premises. It is not necessary to urge a worthy cause on false assumptions. It is very desirable to call attention to bad habits of life, to overeating and drinking, to lack of exercise, all of which result in degenerative processes; and it is certainly to be desired that persons shall have themselves physically examined each year in order to discover impairments and defects when they are only slightly developed rather than to allow them to proceed until a breakdown occurs. This movement deserves the support of all and indeed has given good evidence of excellent returns for the future. In this paper, however, I must insist that the program of life conservation shall not continue to be tied up to or made dependent upon misleading statistics. There is no evidence at hand at the present time to justify the statement that conditions of life in America are increasing the mortality rates at the productive ages.

Digitized by Google

ON THE CORRELATION BETWEEN THE VOTE FOR SUFFRAGE AND THE VOTE ON THE LIQUOR QUESTION. A PRELIMINARY STUDY.

By C. E. Gehlke, Ph.D., *Assistant Professor of Sociology, Western Reserve University.**

In current discussions of the factors tending to the extension of the suffrage to women, and of the effects of that extension, two statements are frequently found. The first is that in the United States, woman suffrage is opposed by those groups in the electorate who are favorable to the continuance of the legalized manufacture and sale of alcoholic liquors. Another is that woman suffrage will be the doom of the legalized sale of liquor, and that fear of this outcome causes the "liquor interests" to oppose suffrage.

This paper presents a preliminary statistical investigation of the probability of the truth of the first statement.

In 1912, the state of Ohio voted on a number of amendments to the State Constitution.† One of these amendments proposed to give the suffrage to women. Another instituted a system of licensing the sale of liquor. Previous to the passage of this latter amendment, the sale of liquor was not licensed but took place under a liquor taxing law.

In 1914, a suffrage amendment was again voted on. At the same election, the question of state-wide prohibition was also presented for referendum action as a proposed amendment to the Constitution.‡

The question of license was not a clear-cut, pro- and anti-liquor issue. It was a matter of common knowledge at that time that many voters preferred prohibition to any system of regulation. Not being able to vote for prohibition, many of them did what seemed to them the next best thing, voted for license. The liquor interests supported the license amend-

* The author wishes to express his obligation to two of his former students, Mr. Fred McNaughton and Mr. C. R. Strobel, for the collection of material for this paper.

† See appendix.

‡ See appendix.

Digitized by Google

ment, for while it put them under regulation, it at the same time legalized the traffic. If license had been a straight out and out pro-liquor proposition the major supposition which we are testing would have given this result: in a given voting district a relatively high vote for suffrage would have been found concomitant with a relatively low vote for license. Expressed in statistical terms, there would have been a high inverse correlation between license and suffrage. If part of the opponents of the liquor traffic had voted for license as giving control, if not elimination, of the sale of alcohol, the degree of inverse correlation would have been lessened compared with what it would have been if a vote had been taken on a clear-cut pro-liquor proposal. To put it quantitatively, if the correlation between the vote for suffrage and the vote for some completely pro-liquor amendment had been, *e. g.*, $-.8$, then the correlation of the pro-suffrage vote and the pro-license vote would have been something less than $-.8$, say $-.5$ or $-.4$.

This general argument applied to the second pair of votes, that of 1914 on suffrage and on prohibition, ought to give a positive correlation for these two; very high, if the sentiment of the anti-suffrage group was strongly in favor of the liquor interests.

The method of testing these probabilities was as follows:

The records of the respective elections in the voting precincts of Cuyahoga County (containing the City of Cleveland, two large suburbs, several smaller ones and a rural and village population) to the number of over five hundred furnished the material. This was arranged on cards, one for each precinct. A sample card is shown.

18—C	For.	Against.	Total Vote.	Per Cent. of Vote in Favor.
(Suffrage)	78	42	120	65.0
(License)	27	94	121	22.3
(Suffrage)	77	33	110	70.0
(Prohibition)	92	41	133	69.1

These five hundred cards were then thoroughly shuffled and there were selected at random two sets of one hundred cards

Digitized by Google

each. Each set would prove a check on the other, serving as a test of the fairness of the choice.

The cards of one set were then sorted for the first correlation (taking suffrage as the subject and license as the relative), into classes as follows: 0.0 to 4.9 per cent., 5.0 to 9.9 per cent., etc. The same class interval was employed for subject and relative. This was repeated with the second set of cards.*

After the correlation tables for suffrage and license were completed, the cards were sorted for the vote on suffrage and prohibition. The suffrage was again the subject and prohibition the relative.†

The similarity between the two tables of each of the elections is marked.

The correlation coefficients were the following:

Table (A)	$r=-.571$	$\pm.046$	Suffrage and license, 1912.
(B)	$r=-.593$	$\pm.044$	
(AA)	$r=+.838$	$\pm.020$	Suffrage and prohibition, 1914.
(BB)	$r=+.894$	$\pm.014$	

These results seem to bear out the supposition which was to be tested. An analysis of the material will make the validity of the method more evident.‡

The total number of votes for and against suffrage in each precinct, was in the large majority of cases numerically greater than the number cast on the question of license.

	Greater.	Equal.	Less.
Table (A)	98	1	1
(B)	97	1	2

The total number of votes on suffrage compared with the total number of prohibition votes showed these results:

	Less.	Equal.	Greater.
Table (AA)	97	0	3
(BB)	99	1	0

* See Tables A and B.

† See Tables AA and BB.

‡ The linearity of table A was tested by the simple criterion cited by William Brown (Mental Measurement, p. 59) from J. Blakeman (*Biometrika*, Vol. IV, p. 349–350)

$$\frac{\sqrt{N}}{.67449} \cdot \tfrac{1}{3}\sqrt{\eta^2-r^2} < 2.5$$

Where η = standard deviation of the means of the arrays divided by the standard deviation of the entire distribution (respectively) of X or Y.

Digitized by Google

If the number of votes in each precinct on suffrage had been equal to the number of votes on license, the correlation between the absolute affirmative votes could have been calculated directly. As they were in nearly every case unequal, the percentage that the affirmative vote on a question was of the total vote cast on that question seemed the fairest measure of the opinion and will of the voters of the precinct.

These facts and this method bring up two possible criticisms which require a brief discussion.

1. It may be objected that these correlation coefficients do not measure the correlation of the two sets of opinions in the mind of an average individual voter. Take a precinct chosen at random:

	For.	Against.	Total.	Per Cent. For.
Suffrage.....	29	103	132	22.0
License.....	108	13	121	89.2

The total number of voters in this precinct was at least 132. At least 11 of them did not vote on the license issue. If there were more than 132 voters, some of the 121 who voted on license, may not have voted on suffrage. If the number of voters had been 132 plus 121, it would have been possible to have had two entirely different groups voting on these two issues.

It is evident then, that the figures we have here are not an accurate statement of the coexistence in the mind of any voter of opinions and desires on these two questions. For if there were 132 plus 121 (*i. e.*, 353) voters in the precinct and 132 voted on suffrage and another 121 voted on license, our data would be absolutely useless for the determination of whether an average individual Z voted for or against suffrage and for or against license. The only way to get at this question would be to note each ballot as it was counted and to make for each voter a table:

Suffrage	No	Yes
License	No	Yes

Digitized by Google

Summing these tables, we should get for a precinct, for those who voted on both issues:

		Suffrage: NO	Suffrage: YES	
License	No	a	b	a+b
	Yes	c	d	c+d
		a+c	b+d	

Where a = the number voting "no" on both issues.
Where b = the number voting "yes" on suffrage and "no" on license, etc.

Such data could be handled by means of Pearson's method for correlation of non-quantitative qualities* or by the association coefficient suggested by Yule.† That would give us an accurate measure of the association of these four desires in the minds of the individual voters.

We have no such data, however. Therefore we must interpret what we have in a somewhat different manner. Let us call the percentage vote in a given precinct for the respective issue a measure of the opinion of the voting body on that issue. If all the voters do not participate it presumes ignorance or a state of indifference regarding this issue on the part of those not voting on it. Our correlation coefficients then express the concomitance of the pairs of opinions of the electorate divided into the more or less fictitious units of voting precincts. In a way, the votes might be said to express neighborhood opinion, since the boundaries of American neighborhoods are so indefinite that the precinct is probably as good a unit as any for our purpose. In short, we may say that our data show that on the average, a neighborhood in Cuyahoga County which was strongly for suffrage was strongly for prohibition and less strongly against license.

* Pearson, K.: Mathematical Contributions to the Theory of Evolution; On the Correlation of Characters not Quantitatively Measurable. Phil. Trans., A, CXCV, pp. 1–47, August 16, 1900.

† Yule, G. U.: An Introduction to the Theory of Statistics, p. 38.

Digitized by Google

2. The use of percentages instead of absolute values raises the question as to the possibility of our correlations being unduly high because of the existence of a "spurious correlation." If the votes were expressed as percentages of the same quantity for each precinct, this objection would hold: "When two characters, a and b are measured in each individual of a series of individuals, and each absolute magnitude is transformed into an index by dividing it by the magnitude of a third character c as found in the same individual, a spurious correlation will be found to exist between the indices of $\frac{a}{c}$ and $\frac{b}{c}$."*

Tested by this statement of Davenport's, there is no spurious correlation here. For, as was pointed out above, the number of votes on suffrage is different from the number cast at the same time for license; similarly for suffrage and prohibition. So that instead of a situation expressed by the symbols $\frac{a}{c}$ and $\frac{b}{c}$ we have, rather $\frac{a(\text{the pro-suffrage vote})}{m(\text{total vote on suffrage})}$, $\frac{b(\text{the pro-license vote})}{n(\text{the total vote on license})}$.

TABLE (A).

PERCENTAGE OF VOTE ON SUFFRAGE CAST IN FAVOR OF SUFFRAGE.

Percentage of vote on license cast in favor of license.	5 | 9.9	10 | 14.9	15 | 19.9	20 | 24.9	25 | 29.9	30 | 34.9	35 | 39.9	40 | 44.9	45 | 49.9	50 | 54.9	55 | 59.9	60 | 64.9	
45–49.9											1		1
50–54.9									1			1	2
55–59.9										1	1		2
60–64.9						2	1	1		1	1		6
65–69.9				2			1		1	2	1	2	9
70–74.9					2	4	1	4	4	7	1	1	24
75–79.9			1	4	2	1	4	1	4	3	1		21
80–84.9			4	2	4	4	1	8			1		24
85–89.9	1	1	4		1	2	1						10
90–94.9			1										1
	1	1	10	8	9	13	9	14	10	14	7	4	100

* Davenport, C. B.: Statistical Methods, p. 54. See also Yule (*op. cit.*), pp. 211–212, and William Brown (*op. cit.*), p. 77.

Digitized by Google

TABLE (B).

PERCENTAGE OF VOTE ON SUFFRAGE CAST IN FAVOR OF SUFFRAGE.

Percentage of vote on license cast in favor of license.	0 | 4.9	5 | 9.9	10 | 14.9	15 | 19.9	20 | 24.9	25 | 29.9	30 | 34.9	35 | 39.9	40 | 44.9	45 | 49.9	50 | 54.9	55 | 59.9	60 | 64.9	65 | 69.9	70 | 74.9		
55–59.9			1									1					2
60–64.9											1	2					3
65–69.9									3	1	3		1	2			10
70–74.9			1			2			2	3	1	1		1	1		13
75–79.9						3	1	6	4	5	1	1					21
80–84.9		1		2	1	4	3	3	3	2	2						21
85–89.9		2	4	5	3	3		2	1	2							22
90–94.9			2		2	1	1			1							7
95–99.9	1																1
	1	3	8	7	6	14	5	11	13	14	8	5	1	3	1	.	100

TABLE (AA).

PERCENTAGE OF VOTE ON SUFFRAGE CAST IN FAVOR OF SUFFRAGE.

Percentage of vote cast on prohibition in favor of prohibition.	5 | 9.9	10 | 14.9	15 | 19.9	20 | 24.9	25 | 29.9	30 | 34.9	35 | 39.9	40 | 44.9	45 | 49.9	50 | 54.9	55 | 59.9	60 | 64.9	65 | 69.9	70 | 74.9	
0–4.9		1													1
5–9.9	2	4				1									7
10–14.9		4	2	1											7
15–19.9			5	6	1										12
20–24.9			1	2	3	1									7
25–29.9			1	1		2	4	1							9
30–34.9				1	2	3	1	1							8
35–39.9				1		4	2	1	1						9
40–44.9							4	6	2	1				1	14
45–49.0			1					1	1	1	2				6
50–54.9									1	8	2				11
55–59.9					1		1			1	1				4
60–64.9												1	1		2
65–69.9												1			1
70–74.9															0
75–79.9										2					2
	2	9	10	12	7	11	12	10	7	11	5	2	1	1	100

Digitized by Google

TABLE (BB).

PERCENTAGE OF VOTE ON SUFFRAGE CAST IN FAVOR OF SUFFRAGE.

Percentage of vote on prohibition cast in favor of prohibition.	5 | 9.9	10 | 14.9	15 | 19.9	20 | 24.9	25 | 29.9	30 | 34.9	35 | 39.9	40 | 44.9	45 | 49.9	50 | 54.9	55 | 59.9	60 | 64.9		
0– 4.9	1													1
5– 9.9	6	4												10
10–14.9	2	2	3	1	2									10
15–19.9		2	3	3	1									9
20–24.9			3	6	3									12
25–29.9				1	4	1		1						7
30–34.9		1			2	1	4	2						10
35–39.9						2		4	1					7
40–44.9							1	5	1	2				9
45–49.9								1	2	6	1			10
50–54.9								1	2	4	4	1		12
55–59.9									1		1	1		3
	9	9	9	11	12	4	5	14	7	12	6	2		100

APPENDIX.

Proposed Amendments to the Constitution of Ohio Voted on at Election of September 3, 1912.

SECTION 1. Every citizen of the United States, of the age of twenty-one years, who shall have been a resident of the state one year next preceding the election, and of the county, township or ward in which he or she resides such time as may be provided by law, shall have the qualifications of an elector and be entitled to vote at all elections.

SECT. 9. License to traffic in intoxicating liquors shall be granted in this state, and license laws operative throughout the state shall be passed with such restrictions and regulations as may be provided by law, and municipal corporations shall be authorized by general laws to provide for the limitation of the number of saloons. Laws shall not be passed authorizing more than one saloon in each township or municipality of less than five hundred population, or more than one saloon for each five hundred population, in other townships and municipalities. Where the traffic is or may be prohibited under laws applying to counties, municipalities, townships, residence districts, or other districts now prescribed by law, the traffic shall not be licensed in any such local sub-division while any prohibitory law is operative therein, and nothing herein contained shall be so construed as to repeal, modify or suspend any such prohibitory laws, or any regulatory laws now in force

Digitized by Google

or hereafter enacted, or to prevent the future enactment, modification or repeal of any prohibitory or regulatory laws. License to traffic in intoxicating liquors shall not be granted to any person who at the time of making application therefor is not a citizen of the United States and of good moral character. License shall not be granted to any applicant who is in any way interested in the business conducted at any other place where intoxicating liquors are sold or kept for sale as a beverage nor shall such license be granted unless the applicant or applicants are the only persons in any way pecuniarily interested in the business for which the license is sought and no other person shall be in any way interested therein during the continuance of the license; if such interest of such person shall appear, the license shall be deemed revoked. If any licensee is more than once convicted for a violation of the laws in force to regulate the traffic in intoxicating liquors, his license shall be deemed revoked, and no license shall thereafter be granted to him. License to traffic in intoxicating liquors shall not be granted unless the place of traffic under such license shall be located in the county in which the person or persons reside whose duty it is to grant such license, or in a county adjoining thereto. The word "saloon" as used in this section is defined to be a place where intoxicating liquors are sold, or kept for sale, as a beverage in quantities less than one gallon.

Proposed Amendments to the Constitution of Ohio Voted on at Election of November 3, 1914.

Article V, Section 1.

SECTION 1. Every citizen of the United States, of the age of twenty-one years, who shall have been a resident of the state one year next preceding the election, and of the county, township, or ward, in which he or she resides, such time as may be provided by law, shall have the qualifications of an elector, and be entitled to vote at all elections.

Article XV, Section 9.

SECTION 9. The sale, manufacture for sale and importation for sale of intoxicating liquor as a beverage is hereby prohibited. The General Assembly shall provide for the enforcement of this provision and enact laws with adequate penalties for the violation thereof.

Digitized by Google

THE PRESENT STATUS OF BIRTH REGISTRATION IN AMERICAN CITIES AND ITS RELATION TO THE INFANT MORTALITY RATE.*

By Louis I. Dublin, Ph.D., *Statistician, Metropolitan Life Insurance Company, New York.*

It requires no extensive statistical research at this time to prove that infant mortality in the United States in recent years has been much reduced and that it can be still further reduced. Communities, however, are not content with this general proposition but desire to know the part they individually play in this reduction. They have their own health problems and they are interested primarily in the facts for their own localities. Cities very properly wish to know what their present infant mortality is, what the trend of that mortality has been during the last five or ten years and what they may hope to accomplish in the next few years in still further reducing their losses. These questions, as we shall see, cannot be readily answered. In this paper I propose to consider the chief limitations on our effort to determine accurately the facts of infant mortality in American cities. I shall show that many of our cities are, indeed, in no position to know the trend of their infant mortality. It must be clear to us that as scientific men we must first insist that our programs to control disease be based on reliable foundations and not on vague impressions or on statistical data subject to very serious error.

The infant mortality rate is the ratio between two figures. The first is the number of births registered in a given period (usually a calendar year); the second is the number of deaths of children under one year of age during the same period, stillbirths being excluded from both figures. There are other measures of infant mortality which are somewhat different; but the infant mortality rate as described is the one in commonest use in present day practice. It is obvious that this infant mortality rate is correct or incorrect in proportion to the

* Paper presented at the Seventh Annual Meeting of the American Association for Study and Prevention of Infant Mortality, Milwaukee, Wisconsin, October, 1916.

Digitized by Google

accuracy of the figures which compose it. If either one or both are wrong the rate will likewise be wrong. Let us consider a few of the possibilities. If the deaths of infants under one are not all recorded, the rate will clearly be too low. Again, if the number of births is, for any reason, not completely registered the infant mortality rate will appear higher than it really is. Finally, if both deaths and births are underregistered, the rate will be wrong except in the rare instances when numerator and denominator are incorrect in the same ratio.

It must be clear, therefore, that the accuracy of the infant mortality rate and of its derivative, the trend of the infant mortality rate, depends on the completeness of registration of births and deaths. Of these two, the registration of deaths has been insisted upon longer by the communities of our country. Considerable state and municipal machinery has been put into operation for this purpose. Laws are in effective operation over a large area of the country, the so-called Registration Area for Deaths, and the interest of the communities and of physicians has been developed to make such registration of deaths more and more complete. We can assume safely that, in the great majority of the cities and states included in the Registration Area, which today comprises over two thirds of the total population of the country, death registration is practically complete. The margin of error is certainly not greater than 10 per cent. and in many cases much less.

Much more serious for our purposes is the matter of birth registration. There is as yet only the beginning of a Registration Area for Births. Adequate legal provision for birth registration exists in a number of states but in many where there is a law it is certain that the law is not enforced. Such is the general consensus of opinion of those most interested in this phase of our vital statistics.

Accordingly, during the course of the last year, we undertook a systematic inquiry into the present status of birth registration throughout the United States. To this end we wrote to the registrars and other officials of all the states and cities which have been included in the Registration Area for deaths since 1910. We asked for the number of births registered in their respective communities during each of the years since

Digitized by Google

1910. In all, about 500 inquiries were sent; replies were received from 16 states and 168 cities. We at once found it necessary to disregard the replies from the state officials; birth reporting is still a local function and it is only with reference to the cities that any inquiry can possibly be worth while. In all, we found only 144 cities whose replies were sufficiently complete to justify analysis. Fortunately, this group of cities is representative of the entire country, since it includes the larger as well as the smaller ones. In the first part of this paper I shall put at your disposal an analysis of the replies from these cities and, upon the basis of the data which they supply, attempt to estimate by a number of simple tests the accuracy and completeness of birth registration in the United States in 1915 and the preceding five years.

The first test of the accuracy and completeness of birth registration is that the number of births registered in any community in a calendar year shall be greater than the number of living children under one year old. For this test the two sets of figures for the year 1910 were available, namely, the census returns for children under one and the replies to our questionnaire giving the number of births registered during the year. We found that the births actually had exceeded the number of children under one in a large number of the cities, although the excess varied considerably from city to city. In 23 out of the 144, or in 15.9 per cent. of all the cities, however, the reverse was true—the population under one did exceed the number of births reported for the year. In all these cities the registration of births was clearly inaccurate and very probably to a high degree; in some, the births registered were certainly less than one half the true number. Table I below gives a list of the 23 cities arranged in the order of the supposed inaccuracy of their birth registration as indicated by this test. The group includes some small towns, but also a number of large cities. Thus we find Baltimore, Chicago, Jersey City, Nashville and Birmingham, Ala., with populations of over 100,000: a few others had populations between 50,000 and 100,000.

Digitized by Google

TABLE I.

CITIES WHERE POPULATION UNDER 1 YEAR OF AGE EXCEEDED BIRTHS REGISTERED IN 1910.

City.	Population under 1 1910.	Excess over Births.	Per Cent. Excess.
Riverside, Cal.	269	9	3.3
Baltimore, Md.	10,239	381	3.7
Winthrop, Mass.	190	9	4.7
Birmingham, Ala.	2,939	162	5.5
Camden, N. J.	2,061	123	5.9
Kearny, N. J.	395	30	7.6
Sunbury, Pa.	361	33	9.1
Norwood, Ohio	297	29	9.8
Fort Wayne, Ind.	1,104	126	11.4
Youngstown, Ohio	1,954	260	13.3
Santa Cruz, Cal.	160	26	16.3
Superior, Wis.	872	153	17.5
Wheeling, W. Va.	781	147	18.8
Elizabeth, N. J.	1,902	362	19.0
Alpena, Mich.	301	64	21.3
Nashville, Tenn.	2,139	470	22.0
Hammond, Ind.	548	133	24.3
Jersey City, N. J.	6,229	1,662	26.7
East Chicago, Ind.	654	232	35.5
Stockton, Cal.	324	135	41.7
Chicago, Ill.	49,073	24,705	50.3
Madison, Wis.	474	257	54.2
Ogden, Utah	647	389	60.1

The second test of the accuracy and completeness of birth registration is the extent to which the birth rate varies from year to year. We should view with suspicion violent changes in the birth rate from year to year or within a few years unless a plausible explanation is at hand. Applying this test we find a number of cities where there has been a decrease between 1910 and 1915. For the most part these decreases are moderate and may very well reflect the true condition of the birth rate in the cities. In some, however, the reduction is very marked indeed. Lynn, Mass., for example, showed a birth rate in 1910 of 26.8 per 1,000; in 1914 the figure had been reduced to 22.3 and in 1915 to 22.1. Such a condition very properly raises a question as to the accuracy of the birth registration in this city in the years 1914 and 1915. In Dover, N. H., the birth rate in 1910 was 23.9; in 1914 it fell to 18.2, and in 1915 rose again to 23.7. These returns also are subject to question. More serious, however, is the fact that in a large number of cities the birth rate increased in a very suspicious manner. In some cities the birth rate in 1915 was close to or even more than twice that in 1910. A list of 8 such cities is given in the following table:

Digitized by Google

TABLE II.

CITIES WITH EXCESSIVE INCREASE IN BIRTH RATES DURING THE PERIOD 1910 TO 1915.

City.	Birth Rate per 1,000 Population.			Per Cent. of Increase, 1910 to 1915.
	1910.	1914.	1915.	
Hammond, Ind.	19.7	33.7	32.4	64
San Diego, Cal.	11.8	22.8	20.4	73
Trenton, N. J.	15.8	31.1	28.0	77
Chicago, Ill.	11.1	21.7	21.1	90
East Chicago, Ind.	21.7	47.1	42.1	94
Stockton, Cal.	8.1	17.0	16.1	99
Ogden, Utah.	10.0	29.8	25.6	156
Madison, Wis.	8.5	25.7	22.4	164

The maximum increase in the birth rate is noted in Madison, Wis., where in 1915 the birth rate was 164 per cent. higher than in 1910. Increases in birth rates as shown in the preceding table are absolutely inexplicable on the basis of a normal increase in fertility and can be explained only on the ground of improved registration. Our conclusion is, therefore, that in the cities listed above the number of births reported in 1910 was entirely untrustworthy and that birth registration had improved perceptibly in 1915. But it must not be concluded from this that the birth registration in 1915 had attained its true value in all of them.

In this connection, an examination of the data for the 144 cities as a whole will be of interest. Because of the belated arrival of some of the 1915 figures we were compelled to compare 1910 with 1914. We found that the number of births in these cities combined increased 20.7 per cent. during the interval between 1910 and 1914. The increase of the population of these cities during the same period was only 10.2 per cent. This last figure must be considered in the light of the fact that during the period the chief source of increase in the population of American cities was through immigration. As might be expected the birth rate for 1914 exceeded that for 1910, the figures being 24.7 and 22.6, respectively. Included in the 144 cities, however, are two of the largest cities in the country, namely, New York and Philadelphia, which together account for 31 per cent. of the total population of the 144 cities. A very different

Digitized by Google

condition is found in them. In New York City the birth rate in 1910 was 26.9 per 1,000 and in 1914, 26.4. In Philadelphia the two figures were 24.9 and 24.8. The two cities combined gave a birth rate of 26.4 in 1910 and 26.0 in 1914. If we eliminate them from consideration in our total we find for the residue that the birth rate has very markedly increased since 1910; it was then 20.9 and in 1914, 24.2. We are accordingly confronted with the interesting fact that in two of the largest cities of the country, where birth registration has been fairly reliable since 1910, the birth rate has decreased; in the 142 other cities, taken as a group, the birth rate has very decidedly increased. In view of the evidence already at hand we are much more likely to be concerned in these figures with an improvement of birth registration in the cities as a unit than with an increase in the birth rate.

To further satisfy ourselves as to this we correlated the increase in the birth rate in the cities between 1910 and 1914 with the condition of the birth rate in 1910. In other words, we attempted to find out by means of a more refined analysis whether a low birth rate in 1910 was in general followed by an increase in the birth rate in the four subsequent years or *vice versa*. Our findings are very clear on this point. We obtained a negative correlation of over .45. This is significant, for it shows that on the whole wherever the birth rate in 1910 was below the average, there appeared during the next four years an increase in the number of births proportionately greater than that in the population and, furthermore, that the greater the deficiency in the birth rate in 1910, the greater the proportionate increase in the birth rate after that year.

There are few who, in view of the results of the tests we have applied, will doubt the incompleteness of birth registration in 1910 or in the next succeeding year or two. We must now consider the more important question whether the birth registration of 1915 may be considered reliable. To this end we arranged our cities in the order of their increasing birth rates for the year 1915. The lowest birth rate in our list of cities was in Santa Cruz, Cal., where the rate was 10.9 per 1,000; the highest birth rate was in Chicopee, Mass., where it was 42.7. The average birth rate was 24.5. If the reader will turn to the

Digitized by Google

table in the appendix in which the birth rates of the 144 cities in 1914 and 1915 are given, he will see that the rates for many of the cities vary markedly from this average. Such variability in birth rates can hardly, I believe, be explained on the basis of climatic, industrial, or racial difference alone. The only conclusion is that many of the cities with extremely low birth rates are still far from having solved their problem of birth registration.

We can at this point apply a third test of the accuracy and completeness of birth registration, the extent to which the birth rate of a city falls below what may be considered as the minimum normal birth rate for American cities. It is difficult, at this time, to lay down a general law as to what the birth rate of American communities should be since that depends upon the constitution of the population and other economic factors. The proportion of foreign to native born is perhaps the most important factor; the age constitution and the proportion of males to females are also to be considered. However, it appears to me to be entirely justifiable to doubt the accuracy of a birth rate in any but a few of our cities which is less than 20 per 1,000 of the population. In some communities where the foreign born stock predominates a birth rate under 25 should be accepted with question. In a number of communities even a birth rate of 25 per 1,000 may be considerably below the truth. For purposes of administrative control, however, a birth rate of 20 should be considered by the health officer as a minimum and his every effort thereafter should be to raise this minimum as his machinery for checking poor registration is improved. In 1915, 32 of the cities included in the study fell below this very conservative minimum. The other tests we have applied, however, indicate that the proportion of the 144 cities with defective birth registration in 1915 is much higher than this.

The defects in present day birth registration must result in very serious embarrassment to the practical worker in the field of infant hygiene, for, as was pointed out at the beginning, incomplete birth registration exaggerates the infant mortality rate and destroys comparability between the figures for different communities. I can illustrate this point with facts for the

Digitized by Google

city of Baltimore in 1910 and thereafter. In 1910, the infant mortality rate for this city was 217.7. In 1915 the rate was 119.8, a reduction of nearly one half. This is a most remarkable showing for so short a period and, if true, it should be a source of great encouragement to the health officers and private agencies of Baltimore concerned with the control of infant mortality in that city. There are, however, a few disturbing facts which must be taken into consideration. Thus in 1910 the recorded birth rate was only 17.6 per 1,000 population; in 1915 it had increased to 23.3. This suggests that we are concerned not alone with a reduction in infant mortality, but also with the effects of improved birth registration. We find, for example, that while the population increased 4.3 per cent. between 1910 and 1915, the number of births actually increased 38.3 per cent. If we apply the birth rate of 1915 to the population of 1910 we would obtain 13,037 births as against 9,858 births which were actually registered. The infant mortality rate for 1910 on this basis would have been 164.6 as against 217.7 which is the rate quoted above. It must be evident, therefore, that the marked decrease in the infant mortality rate is not all clear gain and that a large amount of it is fictitious, being the result of improved municipal bookkeeping.

We could, in like manner, take the returns of a large number of other cities which are undoubtedly proud of their successful campaigns against infant mortality and show, like Baltimore, that much of their supposed life saving is simply a reward for their recent growing interest in birth registration. It would be highly desirable in this connection if the annual circular of the New York Milk Committee on infant mortality rates in the cities of the United States, in singling out cities for praise or blame dependent upon their low or high rates took into consideration the effect of poor birth registration on infant mortality rates. A number of the cities listed in the Milk Committee's annual report for 1915 show exaggerated rates, since the births reported are clearly underestimates. This is especially true for a number of the smaller cities.

In view of the present condition of birth registration and the resulting unreliability of infant mortality rates, what can the statistician offer as a substitute to serve until birth registra-

Digitized by Google

tion in our cities is more complete? Frankly there is no satisfactory substitute for the infant mortality rate. A number of measures have from time to time been used. Thus, the "infant death rate," which is the ratio of the number of deaths of infants to the number of living children under one, has been employed by the Census Bureau. This is, however, subject to very serious error because the number of living children under one is known approximately only for census years and is even then subject to very serious error of misstatement.

Another index is the ratio of the number of deaths of children under one to the total population of all ages as estimated for the year. This measure has the advantage that both numerator and denominator which enter into it can be made fairly accurate. It is especially useful in determining the trend of infant mortality over a short period of years in any one city. However, difficulties at once arise, when it is attempted to compare the rates of different communities with one another. Differences which may appear in such comparisons may not at all be the result of higher or lower actual mortality rates, but rather of the different birth rates in the two places. Thus, if two cities of the same size, say 100,000, have birth rates of 25 and 30, respectively, the city with the higher birth rate will have the larger number of infants born annually, 3,000, as against 2,500 in the second city. A similar infant mortality rate, say of 10 per cent., will result in 300 infant deaths in the one city as against 250 deaths in the other. The actual infant mortality rates would be the same, but the substitute measure will indicate a higher death rate in the city with the higher birth rate, 3.0 as against 2.5 per 1,000 population. Nevertheless for communities where birth registration is known to be very incomplete, the error that results from the use of this substitute index will be less than that resulting from the use of the incorrect number of births. The general principle is that when dealing with factors of error it is safer to work with large numbers than with small ones, for example, with total population than with births.

Another measure which is sometimes used in communities with poor birth registration is the percentage of infant deaths to total deaths. This ratio, like the previous one, has the ad-

Digitized by Google

vantage that the figures entering into it may be presumed to be fairly accurate. The measure, however, has the serious disadvantage of all such proportions in that it does not take into consideration the number of infants exposed. We may very well have, for example, an unusually high proportionate mortality with a low infant mortality rate; this would be the case if the general mortality rate was low Again, a low percentage of infant mortality may occur where the actual infant mortality and the general mortality rates are both high. It is not at all difficult to present a list of cities with high infant mortality rates and low proportions, and conversely.

There is, therefore, no satisfactory substitute for the infant mortality rate. In order to have a measure of the mortality of infants, we must know of necessity the number of infants exposed to death and this means that we must have complete birth registration.

My purpose here is not so much to criticize our present shortcomings as to point out a remedy for a serious evil. Those of us who are interested in the reduction of infant mortality have no choice but to set about to build up machinery for registering officially and quickly every child born. To this end, the way is fairly clear. An adequate law for the reporting of births, the model Law for the Registration of Births, is on the statute books of a number of states. Where it is not, the first step to be taken is to have this law enacted. This would apply to the following states: Alabama, Arizona, Colorado, Delaware, Iowa, Indiana, New Mexico, Nevada, South Dakota, Oklahoma, Texas, and West Virginia. Wherever this law has been or will be enacted it must be enforced. The members of this association can be of the greatest service in this connection either as individuals or collectively to see that every birth in their communities is registered, that as physicians they do this themselves and that as citizens they uphold officials who insist on the fulfillment of the law and the exposure and arraignment of all violators.

In this campaign for the improvement of birth registration the organization with which I am connected desires to play its part. It is probably known to you that for a period of two years, the Metropolitan Life Insurance Company has been

Digitized by Google

engaged in distributing through its agency force thousands of mailing cards, for the purpose of registering births. These cards are addressed to the health officers and registrars of states and cities and are distributed by agents in the homes of policy-holders where there has been a recent birth or where a birth is expected. The mother is directed to fill in the name and date of birth of the child and her own name and address and then to mail the card. As an inducement to the mother to fill out and mail the card, it calls on the health officer to send literature on child hygiene. The form of the card used is shown below.

Metropolitan Life Insurance Company Stat. Form 6 Feb. 1915
NEW YORK

The Health Officer, ...191.....
(DATE)

Dear Sir:

My baby.....
(NAME)

was born on.....191.....
(MONTH) (DAY)

at.....
(NUMBER) (STREET) (CITY)

Please send me your literature on how to keep my baby well.

.....
(PARENT)

The Metropolitan Life Insurance Company takes this means to improve birth registration and to further infant hygiene.

The plan has proved of assistance to many state and municipal health officers. Besides enabling them to register births which otherwise would have escaped official registration, it has, by singling out those who persistently fail in complying with the requirements of the law, given them at least a partial check on delinquent physicians and midwives. If these individuals were followed up with prosecutions and fines, as Dr. Wilbur planned to do in New York State, the number of violations of the law would rapidly be reduced. We have received many communications from registrars all over the country testifying to the aid they have received through our coöperation. I am

Digitized by Google

in a position, at this time, to offer you a more extensive development of this plan. If health officers and other persons interested in the reduction of infant mortality in the several states and cities of the country will write to us, we will see that an adequate supply of the cards is distributed to our policyholders. We will, moreover, follow up this distribution to see that the interest of our field staff is maintained. We are convinced of the efficacy of the plan in improving birth registration.

It is certainly to be hoped that those interested in reducing infant mortality in American cities will concentrate their attention on the problem of birth registration. At the present time we cannot measure with any degree of accuracy, except in a few cities, the effectiveness of our program for the reduction of infant mortality. Too much, as we have seen, is taken for granted and credit is appropriated beyond measure for reductions in infant mortality that are much greater than actually have occurred. We should remove, for all time, the handicap which mars the entire field of our work. Our campaign for the next few years is clearly marked out. We should work to establish a Registration Area for Births which shall be nationwide and as accurate as it is extensive.

Digitized by Google

APPENDIX A.

REGISTERED BIRTHS AND BIRTH RATES RETURNED BY 144 CITIES IN UNITED STATES 1910, 1914, AND 1915.

City.	Number of Births.			Birth Rate per 1,000 Population.		
	1910.	1914.	1915.	1910.	1914.	1915.
Birmingham, Ala.	2,777	3,046	3,040	20.7	18.3	17.5
Mobile	1,058	1,178	1,086	20.5	21.2	19.2
Alameda, Cal.	383	414	442	16.3	15.7	16.4
Los Angeles	5,476	8,308	7,925	16.9	18.9	16.7
Riverside	260	306	278	16.9	16.7	14.6
San Diego	472	1,114	1,045	11.8	22.8	20.4
San Francisco	6,435	7,308	7,649	15.4	16.3	16.8
Santa Cruz	134	162	153	11.9	12.0	10.9
Stockton	189	436	554	8.1	17.0	16.1
Colorado Springs, Colo.	457	432	467	15.6	13.6	14.4
Bridgeport, Conn.	2,984	3,763	3,908	29.1	32.6	33.0
Naugatuck	359	325	361	28.1	23.8	26.0
New Britain	1,607	1,946	1,964	36.3	38.4	37.6
New London	485	617	680	24.6	30.0	32.7
Wilmington, Del.	2,082	2,500	2,600	23.8	27.2	27.9
Washington, D. C.	7,031	7,130	7,067	21.2	20.2	19.7
Jacksonville, Fla.	1,020	1,871	1,688	17.5	26.7	23.1
Savannah, Ga.	1,722	1,744	1,721	26.4	25.7	25.2
Chicago, Ill. (a)	24,368	51,993	51,703	11.1	21.7	21.1
Quincy	879	849	1,000	24.0	23.1	27.2
East Chicago, Ind.	422	1,215	1,145	21.7	47.1	42.1
Fort Wayne	978	1,749	1,564	15.2	24.2	21.0
Hammond	415	824	820	19.7	33.7	32.4
Indianapolis	4,683	5,664	5,417	19.9	21.8	20.4
Logansport	387	406	439	20.3	20.0	21.2
South Bend	1,665	1,906	1,796	30.8	29.3	26.8
Baltimore, Md.	9,858	12,637	13,634	17.6	21.8	23.3
Cumberland	537	709	711	24.5	29.7	27.8
Adams, Mass.	431	484	438	33.0	35.0	31.2
Attleborough Town	426	490	489	26.1	26.8	26.0
Boston	17,670	19,462	19,655	25.7	26.5	26.4
Chicopee	912	1,208	1,207	35.7	43.1	42.7
Haverhill	1,050	1,169	1,214	23.7	24.8	25.4
Lowell	2,650	2,934	2,964	24.9	26.4	26.4
Lynn	2,403	2,186	2,220	26.8	22.3	22.1
Melrose	312	333	354	19.8	19.7	20.6
New Bedford	3,973	3,731	3,673	40.8	33.5	32.0
Newburyport	321	334	367	21.5	22.1	24.2
Newton	826	870	903	20.7	20.5	21.0
Peabody	406	529	588	25.7	30.2	32.8
Quincy	949	1,083	1,075	28.9	29.8	28.9
Revere	462	714	772	25.1	33.1	34.6
Salem	1,260	1,187	1,046	28.7	25.3	21.9
Springfield	2,656	3,226	3,375	29.7	32.1	32.7
Wakefield	276	330	325	24.1	26.8	26.0
Watertown	385	481	500	29.7	33.8	34.4
Webster	404	414	362	34.9	32.7	28.0
Winthrop	181	232	218	17.7	19.6	17.8
Woburn	351	360	369	22.9	22.8	23.3
Worcester	4,060	4,855	4,691	27.7	30.8	29.2
Alpena, Mich.	237	254	215	18.6	19.4	16.3
Detroit	11,509	19,164	20,917	24.5	35.6	37.7
Grand Rapids	2,788	3,205	3,281	24.7	26.0	26.1
Ironwood	478	502	522	37.1	35.5	36.1
Kalamazoo	805	1,022	910	20.2	22.3	19.2
Muskegon	617	716	780	25.6	28.1	30.3
Pontiac	284	453	402	19.4	27.4	23.6
Sault Ste. Marie	353	420	359	27.9	31.9	26.2
Duluth, Minn.	1,877	2,080	2,170	23.6	23.3	23.6
Minneapolis	5,985	7,889	7,813	19.7	23.0	22.1

(a) The number of registered births in Chicago is estimated as 95 per cent. of the returns from Cook County. (See Bulletin 112 of Bureau of the Census, Mortality Statistics for 1911, p. 24, footnote 4.)

Digitized by Google

APPENDIX A—*Continued.*

REGISTERED BIRTHS AND BIRTH RATES RETURNED BY 144 CITIES IN UNITED STATES 1910, 1914, AND 1915.

City.	Number of Births.			Birth Rate per 1,000 Population.		
	1910.	1914.	1915.	1910.	1914.	1915.
St. Paul	3,964	5,195	5,290	18.4	21.9	21.9
Virginia	376	550	488	35.4	40.2	33.8
St. Louis, Mo. (a)	15,663	15,306	15,018	22.7	20.8	20.1
Helena, Mont.	304	383	334	24.2	28.9	24.9
Missoula	242	321	231	18.5	19.5	16.2
Lincoln, Neb.	1,034	1,207	1,169	23.4	26.4	25.4
Berlin, N. H.	470	465	456	39.7	35.7	35.0
Concord	399	402	409	18.5	18.0	18.2
Dover	316	242	314	23.9	18.2	23.7
Laconia	225	270	271	22.0	24.3	24.0
Manchester	2,029	2,242	2,370	28.8	29.6	30.8
Nashua	677	707	719	26.0	26.3	26.5
Asbury Park, N. J.	210	190	222	20.4	14.9	16.6
Bayonne	1,745	2,064	2,287	31.1	31.6	33.8
Camden	1,958	2,444	2,484	20.6	23.9	23.8
East Orange	582	569	784	16.8	14.3	19.0
Elizabeth	1,540	2,196	2,109	20.9	26.6	24.9
Jersey City	4,567	7,258	7,085	17.0	24.7	23.6
Kearney	365	442	486	19.4	20.1	21.4
Montclair	436	491	523	20.1	19.8	20.5
Morristown	250	258	271	19.9	19.8	20.6
Newark	10,289	11,478	11,248	29.4	29.5	28.2
Orange	788	796	791	26.5	24.9	24.3
Passaic	2,040	2,048	1,988	36.9	30.9	28.8
Plainfield	497	569	644	24.1	25.0	27.7
Trenton	1,538	3,327	3,059	15.8	31.1	28.0
West New York	342	597	638	24.9	34.9	35.6
Buffalo, N. Y.	10,008	12,612	12,683	23.5	27.8	27.5
Cortland	265	228	295	22.9	18.1	23.0
Dunkirk	583	569	522	33.6	29.0	25.9
Hornell	227	295	288	16.6	20.6	19.8
Hudson	254	289	377	22.2	23.6	30.1
Little Falls	352	429	488	28.6	33.8	36.8
New York	129,060	140,647	141,256	26.9	26.4	25.8
Niagara Falls	879	1,517	1,434	28.7	43.2	39.6
North Tonawanda	363	419	406	30.2	31.8	30.1
Plattsburg	219	239	227	19.6	19.4	18.1
Saratoga Springs	222	254	266	17.5	19.8	19.3
Schenectady	2,172	2,361	2,165	29.5	26.1	22.7
Syracuse	2,870	3,415	3,579	20.8	22.9	23.5
White Plains	366	545	520	22.7	28.1	25.8
Yonkers	2,107	2,543	2,461	26.2	27.2	25.5
Raleigh, N. C.	463	604	556	24.1	30.5	27.8
Canton, Ohio	1,123	1,278	1,491	22.2	22.8	25.2
Cincinnati	7,263	8,080	7,803	19.9	20.1	19.2
Columbus	3,280	3,837	3,896	18.0	18.8	18.6
Dayton	2,637	2,950	2,697	22.6	23.8	21.5
Massillon	285	278	299	20.5	18.6	19.8
Middletown	306	558	602	23.1	37.6	39.5
Norwood	268	316	247	16.4	15.6	11.6
Sandusky	395	425	418	19.8	21.0	20.7
Toledo	3,542	4,460	4,551	20.9	24.2	24.2
Youngstown	1,694	2,109	2,558	21.2	22.6	24.5
Altoona, Penn.	1,410	1,599	1,525	26.9	28.3	26.5
Carbondale	479	467	446	28.0	25.2	23.6
Erie	1,791	2,300	2,090	26.8	31.8	28.3
Harrisburg	1,634	1,442	1,352	25.4	20.8	19.1
Norristown	571	665	672	20.4	22.0	21.8
Philadelphia	38,676	41,063	40,849	24.9	24.8	24.3
Pittsburgh	15,197	16,328	16,139	28.4	28.9	28.2

(a) April to April.

Digitized by Google

APPENDIX A—*Concluded.*

REGISTERED BIRTHS AND BIRTH RATES RETURNED BY 144 CITIES IN UNITED STATES 1910, 1914, AND 1915.

City.	Number of Births.			Birth Rate per 1,000 Population.		
	1910.	1914.	1915.	1910.	1914.	1915.
Reading	2,496	2,670	2,511	25.9	25.8	23.3
Sunbury	328	377	335	23.7	24.4	21.1
Warren (a)	230	269	297	20.6	19.2	20.6
Pawtucket, R. I.	1,183	1,278	1,419	22.8	22.5	24.4
Providence	5,727	6,098	5,833	25.4	24.9	23.3
Nashville, Tenn.	1,669	2,369	2,095	15.1	20.6	18.1
Galveston, Tex.	685	911	940	18.4	22.6	22.9
Ogden, Utah	258	880	779	10.0	29.8	25.6
Barre, Vt.	247	285	275	22.9	24.3	23.0
Rutland	258	302	322	19.0	20.9	22.0
Richmond, Va.	2,734	3,155	3,473	21.4	23.4	22.5
Seattle, Wash.	4,220	4,905	4,953	17.5	15.7	15.0
Tacoma	1,517	1,546	1,487	17.9	14.9	13.8
Wheeling, W. Va.	634	830	840	15.2	19.4	19.5
Beloit, Wis.	353	415	408	23.2	24.2	23.2
Fond du Lac	431	438	458	22.8	21.5	22.1
Janesville	274	285	287	19.7	20.1	20.1
La Crosse	629	720	621	20.7	23.0	19.7
Madison	217	757	673	8.5	25.7	22.4
Manitowoc	324	296	373	24.8	21.8	27.3
Milwaukee	9,797	11,929	11,278	26.1	28.6	26.3
Oshkosh	750	820	794	22.6	23.4	22.3
Racine	964	1,258	1,350	25.2	28.3	29.7
Superior	719	868	780	17.7	19.6	17.2
Total Cities	466,894	563,678	564,518	22.6	24.7	24.2

(a) Stillbirths included.

Digitized by Google

PROCEEDINGS OF THE SEVENTY-EIGHTH ANNUAL MEETING OF THE AMERICAN STATISTICAL ASSOCIATION, COLUMBUS, OHIO, DECEMBER 27–29, 1916.

The seventy-eighth annual meeting of the American Statistical Association was held at Columbus, Ohio, December 27–29, 1916.

PROGRAM.

WEDNESDAY, DECEMBER 27.

8.00 p. m. First Session, Banquet Hall, Hotel Deshler.

Joint meeting of all the Associations.

Presiding, William Oxley Thompson, of the Ohio State University.

Addresses by the Presidents:

"The National Point of View in Economics," Thomas N. Carver, American Economic Association.

"Countryside and Nation," George E. Vincent, American Sociological Society.

"Opportunities for Practical Work by the Statistical Association," Charles P. Neill, American Statistical Association.

"The Need for Health Insurance," Irving Fisher, American Association for Labor Legislation.

THURSDAY, DECEMBER 28.

10.00 a. m. Second Session, Chapel, First Congregational Church.

1. "Comparative International Statistics," Royal Meeker, U. S. Bureau of Labor Statistics.
2. "Utilization of Statistics in Business," Julius H. Parmelee, Bureau of Railway Economics.
3. "The Increasing Mortality After 45,—Some Causes and Explanations," Louis I. Dublin, Metropolitan Life Insurance Co.

Discussion.

2.30 p. m. Third Session, Chapel, First Congregational Church.

Business Meeting.

10.00 p. m. Smoker, Hotel Deshler.

FRIDAY, DECEMBER 29.

10.00 a. m. Fourth Session, Room 106, General Library Building, Ohio State University.

Joint meeting with the American Economic Association.

"Do the Statistics Regarding the Concentration of Wealth in the United States Mean What They Are Commonly Assumed to,

Digitized by Google

Mean?" Allyn A. Young, Cornell University; Willford I. King, University of Wisconsin.

Discussion: Davis R. Dewey, Massachusetts Institute of Technology; George P. Watkins, Public Service Commission, N. Y. City.

Discussion from the floor under the five-minute rule.

1.00 p. m. A complimentary luncheon by the Ohio State University.

6.30 p. m. Banquet at the Athletic Club, on Broad Street, followed by a discussion of a number of special problems likely to arise at the close of the European War.

Minutes of the Business Meeting.

Meeting called to order at 2.30 p. m., Professor Carl J. West presiding.

The reading of the minutes of the previous annual meeting was dispensed with as they had been published in the Quarterly Publications of the Association.

The Secretary presented the following report:

Secretary's Report.

Mr. Chairman and Members of the Association:

I have the following to report in regard to the condition of the Association:

Membership December 29, 1915, date of last annual meeting	683
New members added	98
Deaths during the year	6
Resignations	22
Dropped	54
Total deductions	82
Present membership	699
Net gain	16
Beside members we have:	
Subscribers (mostly libraries)	215
Domestic Exchanges	40
Foreign Exchanges	117
Total mailing list	1,071

The death of the following members has been reported during the year: E. L. R. Gould, Theodore L. Keppler, E. P. Marshall, W. J. Montgomery, James P. Tolman, Robert G. Valentine.

Digitized by Google

Four regular numbers of the QUARTERLY PUBLICATIONS have been issued during the year, containing an aggregate of 470 pages, or an average of 118 pages, per number.

Only two meetings of the Association have been held during the year; both of these were held in New York City, on May 5 and December 14, respectively. These meetings took the form of a dinner, followed by round-table discussion. The average attendance was about 50.

It is noteworthy that there has been a marked increase in the Association's membership in New York City in recent years, especially since quarterly meetings have been held there rather regularly. It is to be hoped that the members in other large cities will find it worth while to follow the excellent example of our New York members.

Respectfully submitted,

CARROLL W. DOTEN,
Secretary.

On motion, it was voted to accept the report of the Secretary and to authorize its publication in the proceedings in the March number of the QUARTERLY PUBLICATIONS.

TREASURER'S REPORT.

December 23, 1915, to December 22, 1916.

RECEIPTS.

Membership dues	$1,372.15	
Sales and subscriptions	583.96	
Dividends and interest	159.68	
Balance on hand, December 23, 1915	850.97	
		$2,966.76

EXPENDITURES.

Printing	$1,536.36	
Postage	217.88	
Salaries and clerical service	578.46	
Expenses	47.99	
	$2,380.69	
Balance on hand, December 22, 1916	586.07	
		$2,966.76

Digitized by Google

ASSETS.

17 shares B. & A. R. R. stock.

S. B. PEARMAIN,
Treasurer.

AUDITORS' REPORT.

We have audited the accounts of the treasurer of the American Statistical Association, for the year ending December 22, 1916, and counted the securities in his possession. We find the accounts accurately stated and the expenditures properly vouched.

Respectfully submitted,
MARTIN J. SHUGRUE,
ERNEST H. GAUNT,
Auditing Committee.

SEVENTY-FIFTH ANNIVERSARY FUND.

To December 22, 1916.

RECEIPTS.

Cash on hand Jan. 1, 1916.....	$169.36

EXPENDITURES.

July 6. Paid John Koren.....	$87.55	
Nov. 15. Paid John Koren.....	43.75	
Balance on hand December 22, 1916...	38.06	
		$169.36

S. B. PEARMAIN,
Treasurer.

AUDITORS' REPORT.

We have audited the Anniversary Fund accounts of the treasurer of the American Statistical Association for the year ending December 22, 1916, and find the statements correct and the expenditures properly vouched.

Respectfully submitted,
MARTIN J. SHUGRUE,
ERNEST H. GAUNT,
Auditing Committee.

Digitized by Google

On motion, it was voted to accept the reports of the treasurer and the auditors, and to publish them in the proceedings of the meeting.

The report of the Nominating Committee, consisting of W. Z. Ripley, W. S. Gifford, and Fred C. Croxton, which had been sent to the members of the Association thirty days prior to the meeting, in accordance with the provisions of the constitution, was then taken up for action. No other nominations having been made by petition, it was voted to instruct the secretary to cast one ballot for the following persons placed in nomination by the committee: President, Allyn A. Young; vice-presidents, Charles F. Gettemy, Joseph A. Hill, Charles H. Verrill; counsellors, John Koren, E. Dana Durand, Wesley C. Mitchell; treasurer, S. B. Pearmain; secretary, Carroll W. Doten; librarian, Horace G. Wadlin; editor, William B. Bailey. This vote having been cast they were duly declared elected.

Report of Library Committee.

The Library Committee has ascertained with care the existing condition of the Association library, and wishes to make this condition a matter of record by reporting it in full to the Association at this time.

It appears that when the Statistical Association turned its collection of books, pamphlets, and serials over to the Boston Public Library, the transfer took the form of an outright gift. The only condition named was that "the purely statistical portion of the library should be kept together, and members of the Association should have free access to the shelves of the collection." These conditions having been clearly met, it would seem that the gift still stands and that the Association is essentially without any library of its own!

The collection originally transferred numbered about 5,000 books and pamphlets. Some of these were duplicates of copies already in the Public Library. With the consent of the Association these duplicates were eliminated. From material already in the Library, transfers were made to the shelves of the Statistical Department, so that at the end of the year 1903 the Statistical collection numbered 10,756 bound vol-

Digitized by Google

umes. Since that year, additions have been made from three sources: First, gifts transmitted through the Statistical Association; second, exchange of the Association's publications; and third, direct purchase by the Public Library. Upon the average about 875 bound volumes have been shelved annually. Slightly more than half of these appear to have come to the Library through the Association or by direct gift.

It should be noted that not all the books in the collection are distinctly statistical. Works in economics, sociology, and allied subjects are included as a general rule. The Library's policy seems to be to hold within the Statistical Department all books and pamphlets dealing with subjects within the limits of social science. The complete collection numbers today about 20,000 bound volumes.

The collection has been quartered from the first in one of the separate rooms in the rear of the Public Library building. Stacks close beside the tables and in balconies overhead place the books within easy access. But lighting conditions within the room are not what they should be, and the general appearance of the quarters is unattractive. Yet there seems to be no prospect of a better housing of the collection. Naturally a Statistical Department cannot claim either the first attention or best accommodation of a general public library.

Such, then, is the present status of what *used to be* the Association's library. We find ourselves quite unable at this time to frame any definite recommendations looking toward action by the Association.

ROGER W. BABSON, *Chairman.*

After some discussion it was voted that the Library Committee be empowered, after consultation with the authorities of the Boston Public Library, to discontinue any domestic or foreign exchange that is not being utilized by the Library, or to deposit such exchanges in some other library, at the discretion of the committee.

It was further voted to inform the authorities of the Boston Public Library that there is some dissatisfaction on the part of the members of our Association with the way in which the library of this Association, which was placed in their care

Digitized by Google

some years ago, has been utilized by them, and that the Association would be glad to consider the matter of making some other disposition of this collection, in whole or in part, if the trustees are unable to meet our request for more nearly adequate care and a more favorable location for these books and publications.

The report of the Standing Committee on Business Statistics was made by Julius H. Parmelee in the absence of the chairman, W. S. Gifford.

It was moved and carried that the report of this committee be accepted, that the committee be commended for the preliminary bibliography which it has prepared, and that the committee be empowered to continue its work with a view to ultimate publication.

It was voted on motion that the Standing Committee on Federal Statistics be instructed to coöperate with the committees of other organizations in making suggestions to or otherwise assisting the Bureau of the Census in its preparation for the next decennial census.

It was moved and carried that the selection of the place and time for holding the next annual meeting be left with the Board of Directors.

The following resolution presented by Edwin W. Kopf was adopted by unanimous vote:

RESOLUTION.

WHEREAS, The present appropriation law covering expenditures in the several executive departments of the United States Government specially states that no part of the money appropriated shall be expended for paying the expenses of officials or other employees while in attendance at conferences, conventions, or meetings of associations, and

WHEREAS, It is frequently of mutual advantage to the Government and to scientific associations interested in the work of the departments that there shall be at such gatherings free interchanges of views in matters that concern each, and

WHEREAS, The present restriction makes for the curtailing or entire absence of departmental representation at the con-

Digitized by Google

ferences, conventions, or meetings of such associations to the detriment of the public good, be it therefore

Resolved, by the American Statistical Association in annual meeting assembled at Columbus, Ohio, December 28, 1916, That it is the sense of the Association that, whenever in the discretion of the heads of the Government departments or bureaus concerned, it will be to the public interest to send representatives to such conferences, meetings, or conventions, a sufficient amount of the sum appropriated for the maintenance of such departments or bureaus should be available to cover the expenses of such representatives for travel and subsistence during absence from Washington, and be it further

Resolved, That it is the sense of the American Statistical Association that the provision of the present appropriation law which forbids the expenditure of money for the expenses of officials or employees in attendance upon the meetings of scientific associations should be stricken therefrom, and that the secretary be and hereby is instructed to forward copies of this resolution to the President of the Senate, the Speaker of the House, the chairman of the Committee on Finance, the chairman of the Appropriations Committee of the House, and the Secretary of Commerce.

No further business appearing, the meeting adjourned at 4.30 p. m.

CARROLL W. DOTEN,
Secretary.

Digitized by Google

REVIEWS AND NOTES.

The Social Survey, by Carol Aronovici, Ph.D. Philadelphia. The Harper Press. 1916. 255 pp.

This little book, one of the publications of the Bureau for Social Research of the Seybert Institution in Philadelphia, is a welcome addition to the growing literature upon this subject. The first chapter deals with the need of a survey and what may be expected to be accomplished by it. This opening section is, in some ways, the most valuable portion of the book because it was evidently written by one who knows what he is talking about. He recognizes the advantage of skilled assistance and the necessity of limiting a survey to those subjects in which there is some hope for improvement from an awakened public conscience. This section is filled with timely hints gained from experience. The remainder of the book is devoted, to a considerable extent, to schedules which may be used in the making of a survey. These relate to such topics as city planning, industry, health, leisure, education, crime, etc. Under each of these headings is given a list of the particular points upon which information may be expected to be collected. To purely statistical matter, only one section of five pages is devoted and much of this is taken from King's "Elements of Statistical Methods." This very brief reference is little better than nothing. The author excuses the brevity of this treatment on the ground that "All that could be justly expected is to point out the need for its use and refer the reader to the bibliography which gives a list of what are at this time considered the standard works on the statistical method in the hope that the surveying forces will venture into the study of some of these books prior to undertaking the task of tabulating, correlating, and interpreting the information gathered in the course of the survey."

One section of the book gives it such a value that most students of this subject will wish to see this volume in their library. This is a very complete and carefully prepared bibliography of some sixty pages. This is sub-divided under various headings and gives what is by all odds the best list of readings upon this subject which has been prepared up to the present. Taken as a whole, the book is an acceptable contribution to the literature of social problems in this country.

Wm. B. Bailey.

Yale University.

The Forty-Second Annual Report of the Orange Memorial Hospital for the Year 1915.

The annual report of the Orange Memorial Hospital is much like the reports of other hospitals in this country with the exception of the fact that Mr. Frederick L. Hoffman has prepared a statistical analysis of the

Digitized by Google

experience data of the hospital for the year 1915. This analysis enables the reader to gain some accurate information upon a number of points which are not apparent in the ordinary hospital report. The cases are distributed between white and colored and sub-divided according to sex and age. With the cases thus distributed a study is made of the results of the hospital treatment, whether the condition of the patients on discharge was cured, improved, unimproved, died, or not stated. When these distinctions are made it is apparent that the fatality rate was quite different for males than for females and for whites than for colored. The average duration of hospital treatment is also stated according to age and sex.

A further division of the inmates into private, free, or part pay patients is made and these are still further distributed according to the sections of the community from which they come. Although considerably more than half of the free and part pay inmates come from Orange, only about one fourth of the private and semi-private patients come from the same place, showing the advantage to East, West, and South Orange of having a hospital for private patients. This statistical analysis covers but five pages and yet it gives just the information which is often sought in vain concerning the activities of a hospital of this nature. It is to be hoped that other hospitals may be led to include a similar analysis in future reports.

W. B. B.

Select Discussions of Race Problems. A Collection of Papers of Especial Use in the Study of Negro American Problems. Edited by J. A. Bigham. 108 pp. Atlanta University Press. 1916. Price 50 cents.

The first few pages of this pamphlet contain the Proceedings of the Twentieth Annual Conference for Study of Negro Problems. The rest of the volume consists of a number of reprints of articles by well known writers, which have appeared in various publications. The subject-matter of these papers is the physical and mental bases of race distinction, and particularly the question of the superiority or inferiority of races as compared with each other. The particular trend of all the articles is to emphasize the conclusion that no tangible evidence of the inherent racial inferiority of the Negro to the white man, in mind or body, has yet been established. It is obvious that the discussions have been "selected" with this in view, and that therefore the pamphlet, while containing much that is interesting and valuable, is, as a whole, entirely lacking in critical or scientific value.

H. P. Fairchild.

Yale University.

Digitized by Google

SOME STATISTICS RELATING TO BIRTHS IN SOUTH DAKOTA.

The published vital statistics of South Dakota, 1909 to 1913, inclusive, divide the mothers of a given year into two classes according as they give birth in the year to sons or to daughters. Each of these two classes is arranged with respect to increasing age. The same arrangement is followed with respect to the fathers.

The two tables below give some results deduced from these statistics. Ms and Md are abbreviations for mothers of sons and mothers of daughters, respectively. Similarly, for Fs and Fd. The columns headed D and S give the standard deviation and the skewness. The latter is computed with respect to the actual maximum. The column under R gives the range, or the number of years over which births occurred.

The tables are based on more than 63,000 births and covers all the cases for which the appropriate data have been published.

MOTHERS.

Year.	Av. Age Ms.	Av. Age Md.	D. Ms.	D. Md.	S. Ms.	S. Md.	R. Ms.	R. Md.
1909.....	27.9	28.3	6.32	6.37	0.62	0.53	34	35
1910.....	28.5	28.3	6.44	6.45	0.70	0.51	43	40
1911.....	27.8	28.1	6.36	6.35	0.44	0.64	39	37
1912.....	28.3	28.3	6.57	6.36	0.35	0.36	39	35
1913.....	28.2	28.3	6.28	6.37	0.67	0.68	35	36
Mean.....	*28.14*	*28.26*	*6.41*	*6.41*	*0.65*	*0.67*		

FATHERS.

Year.	Fs.	Fd.	Fs.	Fd.	Fs.	Fd.	Fs.	Fd.
1909.....	33.2	33.2	7.66	7.83	0.42	0.41	52	52
1910.....	33.1	34.4	7.72	8.01	0.40	0.42	46	55
1911.....	33.1	32.9	7.72	7.93	0.40	0.37	56	56
1912.....	33.3	32.3	7.74	7.75	0.41	0.41	60	52
1913.....	33.3	32.7	7.60	7.69	0.43	0.35	52	46
Mean.....	*33.2*	*33.1*	*7.69*	*7.84*	*0.418*	*0.392*		

As a partial restatement of the results in the tables we give the following:

In South Dakota, in the years 1909 to 1913, inclusive, the mothers giving birth to daughters showed less variability in age, in the distribution of age with respect to the yearly average and in the assymmetry of their distribution than did mothers giving birth to sons.

In South Dakota, in the years 1909 to 1913, inclusive, the fathers of sons born in this period were less variable in average age, in the distribution of ages with respect to the yearly average, and in the assymmetry of this distribution than did the fathers of daughters.

Taking the entire period into consideration the average age of mothers of daughters differs less from that of the fathers than does the average age of the mothers of sons differ from that of the fathers.

Digitized by Google

The average age of the mothers of daughters is greater than that of the mothers of sons and the average age of the fathers of daughters is less than that of the fathers of sons.

T. E. M.

Vermilion, S. D.

"THE HOMICIDE RECORD OF AMERICAN CITIES FOR 1915," THE SPECTATOR, DECEMBER 21, 1916. F. L. HOFFMAN.

Many agencies, both public and private, are indebted to Mr. Hoffman for his annual review of the homicide record in American cities. There is no other source to which one may resort for a general compilation and analysis of statistics relating to this particular crime. True, the Division of Vital Statistics of the United States Census Bureau keeps a record of mortality data throughout the registration area of the United States, but this division is considering homicide merely as one of the many causes of deaths and not as a crime. Mr. Hoffman has used these data as a basis for a study of the amount of homicide considered as a crime. That the only effort to present this branch of criminal statistics for the country as a whole is left to be undertaken year after year by a single individual in unofficial life is no less a tribute to Mr. Hoffman than it is a condemnation of the various agencies of government which should be concerned with accumulating and interpreting this information. It is to be hoped that while Mr. Hoffman continues to fill the breach, by his presentation of this branch of criminal statistics, a stimulus may be given to other agencies, particularly governmental, to appreciate the value of such record information, and to undertake the preparation of this and other branches of criminal statistics on a comprehensive scale.

Mr. Hoffman has used the scattered returns of prison statistics prepared by the Census Bureau and the mortality returns of the Division of Vital Statistics of the Census Bureau and of municipal departments of health. The prison statistics of 1890 and 1910 relate to a point of time, showing the number of prisoners in confinement on a given day. But while the statistics of these enumerations speak a point of time, they really deal with a period of time, since the prisoners were brought into confinement on separate occasions, ranging over a long period. This period is indefinite and unknown. The Census Bureau investigation of June 1, 1890, shows that there were 7,386 prisoners in the United States charged with homicide, or 9 per cent. of the total number of prisoners under confinement; and again, a census return of January 1, 1910, shows 6,904 prisoners who were charged with first degree homicide, and 7,412 for second degree homicide, making a total of 14,316. By preparing a population ratio, Mr. Hoffman shows that the number of prisoners held on these charges was 156 per 1,000,000 inhabitants in 1910, as against 117 per 1,000,000 in 1890. On this basis the assumption is made that "in the direction of the security of the person there has been decided retrogression." For the purpose of

Digitized by Google

determining the amount and nature of this particular phase of criminality, the value of these totals becomes extremely doubtful, since the figures are acquired as of a point of time yet applying to many periods of time, which range from one day to twenty years or over, according to the length of time the prisoner has served his sentence. Moreover, the probability is strong that a number of the prisoners charged with homicide, involving life sentences, or twenty-year terms, who were counted in the enumeration of June 1, 1890, were counted again in January 1, 1910. The per centum calculation of the proportion of prisoners charged with homicide to the total number of prisoners is likewise misleading. To say that in 1910, 10.5 per cent. of the total number of prisoners, as against 9 per cent. in 1890, were charged with homicide makes it appear that one tenth of all persons convicted had committed homicide. It must be pointed out that the persons imprisoned on a charge imposing long sentences have been accumulating throughout the years of a long period of time, whereas there are many charged with lesser offenses, such as burglary, grand larceny, etc., who have been imprisoned during the same period, and whose terms have expired before the day on which the enumeration was made.

The prison statistics are used only to point the general status of homicide —mortality returns are used as the "most trustworthy" data on the subject of homicide. The point is indeed well taken that there is a very appreciable difference between the number of victims of homicide and the number of persons convicted of homicide. For this reason, mortality statistics are better than judicial statistics in assessing the total amount of homicide criminality. Mortality data are reliable since a death is a non-varying unit, but after all, mortality statistics are not crime statistics, and the chief consideration of a homicidal death should be its criminal aspect. A study of homicide as a crime must take into account other elements than the victim and the manner of inducing death. The perpetrator of a homicide is the most important factor. The motive of the crime is perhaps more important than the manner of commission. Then there are several degrees of homicide that must be reckoned with. Since judicial statistics do not include all the crimes committed, and mortality statistics do not include all the facts essential to a study of homicide as a phase of criminality, there is imperative need that we look for some other source that will meet all the requirements. It is possible that judicial statistics can be improved to the extent that they include the elements now lacking. It may be well here to point out the possibilities of police statistics. The police department is an agency which must take account of all homicides that are committed within its jurisdiction, and record the data relating to both victim and perpetrator, the manner of commission and the motive. Finally, the police are also concerned with the judicial phase of cases in which arrests and indictments are made. Considerable progress has already been made by some police departments in this connection, notably New York and St. Louis.

The conclusion that the single case of murder of a commissioner of banking by bank officials "is of special interest as emphasizing the lament-

Digitized by Google

able fact that the homicidal impulse is not by any means limited to the lower classes or to foreigners of recent immigration to this country," while correct as far as absolute limitation goes is not sufficiently supported by facts to admit of the conclusion implied. A comprehensive study of the perpetrators of homicide is needed to support intelligent conclusions as to what racial or social groups are responsible for murders and manslaughters. It is necessary to make such an analysis according to cities and other areas in much the same way that the mortality data has been considered. Such a study would doubtless show that the Negro is not alone the victim of homicide in the South, as Mr. Hoffman points out, but that he is also the perpetrator, and responsible for a disproportionate number of homicides, according to the Negro population, in certain Northern cities. For example, the police statistics of Chicago show that there were 33 Negro decedents and 32 Negro perpetrators in murder and manslaughter cases for the year 1915. In St. Louis there were, in 1915, 20 Negro decedents in murder and manslaughter cases and 24 perpetrators known to have been Negroes, out of a total of 71 decedents and 72 perpetrators. In Washington, D. C., in 1916, there were 18 colored decedents, as against 6 white, and 18 colored perpetrators, as against 5 white perpetrators.

There is some disparity in the figures for homicide which Mr. Hoffman presents in "Table V—Mortality From Homicide in Thirty-one American Cities, 1905–1915," and the figures for murder and manslaughter which I have obtained from the police departments of certain cities for the year 1915, as shown below.

	Mortality Table.	Police Records.
Memphis	125	75
St. Louis	103	74
Chicago	211	198
New York	258	247
New Orleans	185	71

The differences in Chicago, New York, and New Orleans may be accounted for in the fact that the mortality figures include abortions, while the police figures do not. This same fact may account for the difference of from 10 to 15 in St. Louis and Memphis. The remaining difference in the last named cities may be due to the fact that the mortality returns include homicidal deaths in city hospitals or other places in the city which resulted from crimes committed outside the limits of the city. The police figures deal with only those crimes committed within the city limits.

The tables showing the geographical incidence of homicide by states and cities are particularly illuminating. Altogether the work of Mr. Hoffman is very creditable and as accurate as can be expected under the present haphazard system employed in the United States in the reporting of homicides.

Raymond B. Fosdick.

New York City.

Digitized by Google

STATISTICS OF RELIGIOUS FAITHS IN NEW YORK CITY.

On Sunday, February 4, a body of 34,000 enumerators acting under the general direction of Mr. J. Shreve Dunham, ascertained the religious preferences of nearly three quarters of New York City's population. The city was districted and the information was secured through a thorough house to house campaign. The final program was approved at a meeting of priests, ministers, and rabbis representing the congregations of the several churches and religious societies in the city. Mr. Dunham will shortly go to Denver, Col., to conduct a similar campaign.

HISTORY OF THE FEDERAL REGISTRATION SERVICE.

In a monograph "The Federal Registration Service of the United States: Its Development, Problems, and Defects," Dr. Cressy L. Wilbur traces the work of the United States Government in vital statistics since the Seventh Census in 1850 and up to the close of the year 1915. The report pays particular attention to the continuous registration of deaths in specified areas since the "registration area for deaths" was established in 1880. Comprehensive tables and cartograms exhibit the main facts of the movement for satisfactory legislation and the enforcement of legislation. The section on "When and How Will the United States Obtain Complete Registration of Vital Statistics?" is practically a re-statement of Dr. Wilbur's prophetic conclusions in the paper which he read before the American Public Health Association at Denver, in 1895, and a summary of vital statistics progress of the last two decades. The document is issued by the Bureau of the Census, Washington, D. C.

NOMENCLATURE AND CLASSIFICATION OF CAUSES OF DEATH.

A Committee of the Vital Statistics Section of the American Public Health Association on the accuracy of certified causes of death has issued a summary of its conclusions following a detailed study of each title in the International List of Causes of Sickness and Death and of the terms included under such titles. The committee considered (a) the validity of the terms included under each title, *i. e.*, the probability that each term when reported as a cause of death represents the exact morbid condition covered by the title with which it is classified, and (b) the status of the term as descriptive of the *primary* cause of death. It was also the aim of the committee to separate the titles of the International List into two classes: (a) Those not to be accepted as reliable without autopsy or without specific supporting data upon which the diagnosis was based and (b) Those to be accepted as reliable without autopsy or other verification. The report of this committee is a further contribution to the critical literature on statistics of causes of death. Dr. Haven Emerson, Chairman, and Mr. George H. Van Buren, Executive Secretary, prepared the summary as published

Digitized by Google

by the United States Public Health Service under Reprint Number 363. A useful bibliography of this subject, including a reference to the Oertel-Bashford-Hoffman articles on the accuracy of cancer mortality statistics in the London *Lancet*, is to be found on p. 3 of Mr. F. L. Hoffman's "Mortality from Cancer Throughout the World."

STANDARDS IN STATISTICS OF INDUSTRIAL ACCIDENTS AND AN ADVANCE STEP IN THE CLASSIFICATION OF INDUSTRIES.

The Committee on Statistics and Compensation Insurance Cost of the International Association of Industrial Accident Boards and Commission has published its report as Bulletin 201, United States Bureau of Labor Statistics. A preliminary report had been previously published in the *Monthly Review* of the bureau for November, 1915. The classification of accident causes recognizes twelve main classes of the general subject. Each of these general classes is suitably partitioned in subordinate ranks. A supplementary classification of location and nature of injury and extent of disability and degree of partial disability is appended. The grouping of working machinery employed by the New York State Industrial Commission is published as Appendix A.

The classification of industries is a decided advance over any available since the publication of the United States Census Bureau Index to Occupations in 1910. In plans for the classification of occupations in the Census of 1920, the I. A. I. A. B. C. industry scheme will probably serve as a better base than the industry outline published in Volume IV, Thirteenth Census of the United States. The statistics published in that volume seem not to have taken the industry scheme of the 1910 Index and of the Bertillon-Vannacque International Statistical Institute List into account. For the purposes of most social statistics an industry classification of occupations is essential.

E. W. Kopf.

New York City.

Income Tax Procedure. By Robert H. Montgomery. The Ronald Press. New York. 1917. Pp. X, 461. $2.50 postpaid.

There is nothing statistical about this book, but in view of the fact that the American Statistical Association adopted a resolution at its annual meeting in 1914, calling upon the Bureau of Internal Revenue to so modify its income tax blank as to make it possible to obtain accurate statistics of incomes in this country, and in view of recent conferences between that bureau and a joint committee of this Association and the American Economic Association on this same matter, it may be well to direct the attention of statisticians to this work. The author disclaims any intention of presenting a complete digest of the income tax law. He says also that his

Digitized by Google

work is not a treatise or a history of the subject, though he does treat these aspects very briefly in a short historical chapter.

The real purpose of the book is to answer for the perplexed taxpayer a large number of questions which arise in his mind when he undertakes to make his returns. It explains the duties of those who must make returns, the kinds of income that are taxable, the various exemptions and deductions allowed, and the method of procedure in the matter of protests, overpayments, rebates, appeals, etc. All of this instruction and information is based upon a careful study of the rulings and decisions of the U. S. Treasury Department.

The book also contains chapters on state and municipal income taxes, the "Corporation Excise Tax," and the "Munitions Manufacturers' Tax."

The author has not, however, confined his efforts to description and exposition alone. He has not only pointed out the defects in the law and criticized its administration, but he has offered constructive suggestions for improvement in a number of instances.

It would have been an aid to the reader if the full text of the law had been printed in one place, even though it would have added somewhat to the size of the book, but this defect is not a serious one in view of the ease of obtaining official copies of the law.

On the whole the book is an acceptable and useful guide to the student as well as to the tax payer.

CARROLL W. DOTEN.

Massachusetts Institute of Technology.

Digitized by Google

Publications of the American Statistical Association.

The New Series of the Association, issued quarterly, will be sent to members of the Association on payment of the annual dues of two dollars. The subscription price is also two dollars per annum. Separate copies can also be purchased upon remittance of the advertised prices.

The following list of titles includes only the most important papers published in the several issues. There is in addition a considerable amount of miscellaneous material.

VOL. I. 1888-1889. Nos. 1-8. Pp. 492. PRICE $4.00.

No. 1. Statistics of Water Power Employed in Manufacturing in the United States. GEORGE F. SWAIN. 75 cents.

Nos. 2, 3. Park Areas and Open-Air Spaces in American Cities. E. R. L. GOULD.

Key to the Publications of the United States Census, 1790-1887; with Occasional References to Other Statistical Works. EDWARD CLARK LUNT. $1.00.

No. 4. Life Insurance in the United States. WALTER C. WRIGHT. $1.00.

No. 5. Notes on the Statistical Determination of the Causes of Poverty. AMOS G. WARNER. REMARKS. SAMUEL W. DIKE.

Statistics of Divorce in the United States and Europe. SAMUEL W. DIKE. 50 cents.

No. 6. American Railroad Statistics. ARTHUR T. HADLEY.

Statistics of Municipal Finance. HENRY B. GARDNER. 75 cents.

No. 7. Prison Statistics of the United States for 1888. ROLAND P. FALKNER. 50 cents.

No. 8. Finance Statistics of the American Commonwealths. EDWIN R. A. SELIGMAN.

Divorce in France. BENJAMIN F. KELLER.

Relief of the Poor in Germany. A. G. WARNER. $1.00.

VOL. II. 1890-1891. Nos. 9-16. Pp. 470. PRICE $4.00.

No. 9. Mortgage Statistics. GEORGE K. HOLMES. 50 cents.

No. 10. The Study of Statistics in Italian Universities. CARROLL D. WRIGHT.

Statistics of Private Corporations. ROLAND P. FALKNER. 50 cents.

Nos. 11, 12. Statistics of the Colored Race in the United States. FRANCIS A. WALKER.

How to Make Statistics Popular. CHARLES F. PIDGIN.

Retail Prices in Boston and Vicinity. WALTER F. COOK. $1.00.

No. 13. The Growth of Cities in Massachusetts. HORACE G. WADLIN.

Rate of Natural Increase of Population in United States. HERMAN HOLLERITH.

The First Census of Massachusetts. SAMUEL A. GREEN.

The Commercial Death Rate. ALBERT C. STEVENS. $1.00.

No. 14. A Basis for Statistics of Cost of Production. CARROLL D. WRIGHT.

A Formula for Predicting the Population of the United States. H. S. PRITCHETT.

Weight and Longevity. T. B. MACAULAY.

Some Results of Sanitary Legislation in England since 1875. GARY N. CALKINS. 75 cents.

Digitized by Google

No. 15. The Eleventh Census. R. P. PORTER.
Criminal Statistics. ROLAND P. FALKNER. 75 cents.

No. 16. The Census Enumeration in Prussia. CARL C. PLEHN.
A Plea for the Average. GEORGE K. HOLMES. 75 cents.

VOL. III. 1892-1893. Nos. 17-24. Pp. 614. PRICE $4.00.

No. 17. Statistics of Crime in Massachusetts. FRED G. PETTIGROVE.
Development of Statistics of Religions. M. FOURNIER DE FLAIX. TRANSLATED BY ALICE R. JACKSON.
Net Profits of Manufacturing Industries in the State of Massachusetts. FREDERICK B. HAWLEY.
Classification of Trade Statistics. FREDERICK C. HICKS. 75 cents.

Nos. 18, 19. A Statistical Inquiry concerning Domestic Service. LUCY M. SALMON.
The Theory and Practice of Price Statistics. R. P. FALKNER.
Measures of Distribution. GEORGE K. HOLMES.
Statistics of Suicides in New England. DAVIS R. DEWEY.
Semi-Annual Census of Paupers in Minnesota. H. H. HART. 75 cents.

No. 20. The Establishment of an Imperial Commission for Labor Statistics in Germany. CARL C. PLEHN.
French Statistical Publications. CHARLES D. HAZEN.
Sketch of Life of Robert W. Wood, M. D. JOHN WARD DEAN. 75 cents.

Nos. 21, 22. Observations on the Measure of Change. C. H. COOLEY.
Cost Statistics of Public Electric Lighting. VICTOR ROSEWATER.
Immigration and the Foreign Born Population. RICHMOND MAYO-SMITH.
The Value of Percentile Grades. LUTHER GULICK. $1.00.

No. 23. Murders in Massachusetts. WALDO M. COOK.
Classification of Occupations in the Census. TRANSLATION OF REPORT MADE BY M. BERTILLON.
Growth of Cities in the United States during the Decade 1880-90. CARL BOYD.
The Vital Statistics of an Apache Indian Community. 75 cents.

No. 24. Statistical Data for the Study of the Assimilation of Races and Nationalities in the United States. RICHMOND MAYO-SMITH.
Report of an International Mortality Standard, or Mortality Index. JOSEPH KÖRÖSI.
Character and Volume of the Money of the United States. MAURICE L. MUHLEMAN.
Fluctuations in the Secured Circulation of the National Banks and their Relations to the Prices and Investment Values of Bonds. CHARLES A. CONANT.
Currency Reform in Austria-Hungary. DR. JULIUS MANDELLO.
Comparative Statistics of Primary Education. E. LAVASSEUR.
Results of Recent Investigations on Prices in the United States. F. W. TAUSSIG.
Geographical Concentration: an Historic Feature of American Agriculture. JOHN HYDE.
The Course of Wages in the United States since 1840. CARROLL D. WRIGHT.
Some Recent Results in Railway Statistics in the United States. HENRY C. ADAMS.
The Condition and Needs of Statistics of Marriage and Divorce. SAMUEL W. DIKE.

Digitized by Google

Railway Statistics as Applicable to Earnings of Passenger Trains M. RIEBENACK.
Comparability of Trade Statistics of Various Countries. A. E. BATEMAN.
The Geographical Distribution of the Population of the United States. HENRY GANNETT.
A Preliminary Report on Anthropometry in the United States. EDWARD MUSSEY HARTWELL.
Remarks on the Theory of Anthropometry. FRANZ BOAZ.
On the Application to Individual School Children of the Mean Values Derived from Anthropological Measurements by the Generalizing Method. W. TOWNSEND PORTER.
Anthropometric Statistics of Amherst College. EDWARD HITCHCOCK.
An Anthropometrical Study of the Effects of Gymnastic Training on American Women. CLAËS J. ENEBUSKE.
Railway Freight Traffic Statistics. C. P. LELAND. $1.50.

VOL. IV. 1894-1895. Nos. 25-32. Pp. 387. Price $4.00.

Nos. 25, 26. The Marriage Rate in Michigan. 1870-90. WALTER F. WILCOX.
The Classification of Occupations for Census Purposes. HENRY GANNETT.
Nativity and Occupation of Members of the Massachusetts Legislature. F. H. HOWLAND. 75 cents.
No. 27. Causes of Poverty Further Considered. AMOS G. WARNER.
Congress of Hygiene and Demography. R. P. F. 50 cents.
Nos. 28, 29. Computation of the Value of the Wealth in Existence. GEORGE K. HOLMES.
Effects of Free Surrender and Loan Privileges in Life Insurance. MILES MENANDER DAWSON.
Was the Count of Population in 1890 Reasonably Correct? HENRY GANNETT. 75 cents.
No. 30. Receipts and Expenditures of Certain Wage-Earners in the Garment Trades. ISABEL EATON.
The Negro in the West Indies. FREDERICK L. HOFFMAN. 75 cents.
No. 31. Almshouse Women. MARY ROBERTS SMITH. 75 cents.
No. 32. Accidents in Factories and Elsewhere. KATHARINE PEARSON WOODS.
The Marriage Rate in Massachusetts. F. S. CRUM. 75 cents.

VOL. V. 1896-1897. Nos. 33-40. Pp. 389. PRICE $4.00.

No. 33. Real Estate Values in Boston. HENRY WHITMORE.
Ethnic Influences in Vital Statistics. W. Z. RIPLEY. 75 cents.
No. 34. Railway Competition: A Problem in Statistics. H. T. NEWCOMB.
Mortality in Twenty-Three Massachusetts Cities. CHARLES E. BURNAP.
Note on Methods of Estimating Population. C. L. WILBUR. 50 cents.
Nos. 35, 36. Round Numbers in Wages and Prices. EDWARD D. JONES. $1.00.
No. 37. Remarks of President Walker at Washington.
Vital Statistics for the Twelfth United States Census. CRESSY L. WILBUR.
The Vital Statistics of the Census. WILLIAM A. KING.
General Francis A. Walker. 75 cents.

Digitized by Google

No. 38. Francis Amasa Walker. ADDRESS BY CARROLL D. WRIGHT.
Bibliography of the Writings and Reported Addresses of Francis A. Walker. 50 cents.

No. 39. Some Characteristics of Farm and Home Proprietors. GEORGE K. HOLMES.
A Trial Bibliography of the Writings of Johann Peter Süssmilch, 1707–1767. W. F. WILLCOX AND F. S. CRUM. 50 cents.

No. 40. Farm Ownership and Tenancy in the United States. EDWARD ATKINSON AND L. G. POWERS.
Vital Statistics at the American Public Health Association. 50 cents.

VOL. VI. 1898-1899. Nos. 41-48. Pp. 398. PRICE $4.00.

No. 41. Obstacles to Accurate Statistics. JAMES H. BLODGETT.
Uniformity in Census Returns. JOSEPH KÖRÖSI.
Cost of Production of Corn and Oats in Illinois in 1896. NATHAN A. WESTON. 50 cents.

No. 42. Concentration in Pig-Iron and Coal Production. E. W. PARKER.
The Annual Statistics of Manufactures in Massachusetts. HORACE G. WADLIN. 50 cents.

No. 43. Comparative Statistics of Railroad Rates. HENRY H. SWAIN
Comparative Statistics of Railroad Service under Different Kinds of Control. C. E. PREVEY. 50 cents.

No. 44. Considerations in Gathering Forestry Statistics. B. E. FERNOW. 50 cents.

No. 45. Contributions to the Study of Wage Statistics. CHARLES J. BULLOCK.
A Year of State Deficits. WORTHINGTON C. FORD. 50 cents.

No. 46. The Growth of the Population of Boston. FREDERICK A. BUSHEE.
Wage Statistics in Theory and Practice. ROLAND P. FALKNER.
A Comparative Study of the Statistics of Agriculture of the Tenth and Eleventh Census. N. I. STONE. 75 cents.

No. 47. Notes on Map Making and Graphic Representation. W. Z. RIPLEY.
The Portuguese Population in the United States. FREDERICK L. HOFFMAN.
Report on Uniform Financial School Reports. 50 cents.

No. 48. Development of the Plan for a Census of the World. JOHN HOWARD DYNES. 50 cents.

VOL. VII. 1900-1901. Nos. 49-56. Pp. 466. PRICE $4.00.

Nos. 49, 50. Statistics of College and Non-College Women. MARY ROBERTS SMITH.
The Comparative Accuracy of Different Forms of Quinquennial Age Groups. ALLYN A. YOUNG.
Aids in the Use of Government Publications. LUCIUS PAGE LANE. 75 cents.

No. 51. The Registration Laws in the Colonies of Massachusetts Bay and New Plymouth. ROBERT RENÉ KUCZYNSKI.
Tramps and Wayfarers. 50 cents.

No. 52. Advertising in the United States. SIDNEY A. SHERMAN.
The Practical Use of Vital Statistics. FREDERICK L. HOFFMAN. 75 cents.

Digitized by Google

No. 53. Industrial Consolidations in the United States. LUTHER CONANT, JR.
The Enumeration of Children. ALLYN A. YOUNG. 75 cents.

No. 54. Some Statistics of Recidivism among Misdemeanants in Boston. JOHN KOREN. 75 cents.

No. 55. The Statistical Work of Süssmilch. F. S. CRUM. 75 cents.

No. 56. Permanent Census Office. 75 cents.

VOL. VIII. 1902-1903. Nos. 57-64. Pp. 422. PRICE $4.00.

No. 57. Notes on the Fraternal Beneficiary Corporations Doing Business in Massachusetts. FRANCIS B. FORBES. 50 cents.

No. 58. What is the Centre of an Area, or the Centre of a Population? JOHN F. HAYFORD.
Element of Labor in Railway Expenditure. ANDREW L. HORST. 75 cents.

No. 59. A Statistical Study of the Fatality of Typhoid Fever at Different Seasons. C.-E. A. WINSLOW. 50 cents.

No. 60. The Vital Statistics of the Census of 1900. FREDERICK L. HOFFMAN. 75 cents.

No. 61. The Negroes of St. Louis. LILLIAN BRANDT. 75 cents.

No. 62. A Statistical Study of Patients admitted at the Connecticut Hospital for Insane from the Years 1868 to 1901. ROLLIN H. BURR. 50 cents.

Nos. 63, 64. Urban and Rural New England. WILLIAM B. BAILEY. $1.00.

VOL. IX. 1904-1905. Nos. 65-72. Pp. 356. PRICE $4.00.

No. 65. The Decrease of Consumption in New England. DR. S. W. ABBOTT.
Index Numbers of Prices. 50 cents.

No. 66. The Negro as a Peasant Farmer. KATHARINE COMAN.
Census Statistics of Special Classes. JOHN KOREN. 50 cents.

Nos. 67, 68. The Bread and Meat of the World. YVES GUYOT. $1.00.

No. 69. Recent Railway Accidents in the United States. CARROLL W. DOTEN. 75 cents.

No. 70. Methods of Measuring the Concentration of Wealth. M. O. LORENZ.
Homicide in New Hampshire. HARRY G. NUTT. 75 cents.

No. 71. The Birth-rate in New Hampshire. ALLYN A. YOUNG.
The Growth of New York State Census. J. H. MIDDLETON.
The Results of the Practical Abolition of Capital Punishment in Belgium. MAYNARD SHIPLEY. 75 cents.

No. 72. Methods of Presenting Statistics of Wages. W. C. MITCHELL. 50 cents.

VOL. X. 1906-1907. Nos. 73-80. Pp. 538. PRICE $4.00.

No. 73. The General Death-rate of Large American Cities, 1871-1904. FREDERICK L. HOFFMAN. $1.00.

No. 74. The World's Recent Production of Gold and its Influence upon Commodity Prices. FRANCIS B. FORBES.
A New Method of Index-numbers for American Commodity Prices. FRANCIS B. FORBES. 75 cents.

Digitized by Google

No. 75. Death-rate of the United States in 1900. WALTER F. WILLCOX. Methods of Dealing with Birth-rate Statistics. ED. CARROLL W. DOTEN. 75 cents.

No. 76. The Determination of Racial Stock Among American Immigrants. RICHARD BOECKH. 50 cents.

No. 77. Some Statistical Problems connected with the World's Stock of Gold. FRANCIS B. FORBES.
The Massachusetts Probation System: Its Administration and Operation.
Homicide and the Death Penalty in Austria-Hungary. MAYNARD SHIPLEY.
Statistics of Child Suicide. ARTHUR MACDONALD. 50 cents.

No. 78. Physical Welfare of School Children. $1.00.

No. 79. The Dewey Report on Wages in Manufacturing Industries in the United States. A. E. JAMES.
The Upper East Side: A Study in Living Conditions and Migration. MARY LOUISE MARK.
Some Facts about the Prison Population in the United States. JOHN KOREN. 75 cents.

No. 80. Vital Statistics: Fifteen Papers read at the Meeting of the Section on Vital Statistics of the American Public Health Association at Atlantic City, October, 1907. $1.00.

VOL. XI. 1908-1909. Nos. 81-88. Pp. 753. PRICE $4.00.

No. 81. Address of Carroll D. Wright, President of the American Statistical Association, at its Annual Meeting in Boston, Jan. 17, 1908.
The Outlook for Statistical Science in the United States. S. N. D. NORTH.
An Interpretation of Certain Statistical Evidence of Concentration of Wealth. G. P. WATKINS.
The Measurement of Social Pressure. FRANKLIN H. GIDDINGS.
The Bill for the Thirteenth Census. WALTER F. WILLCOX. $1.00.

No. 82. Problems of Social Statistics and Social Research. FREDERICK L. HOFFMAN.
City Life and Male Mortality. J. E. BAKER.
The Use and Misuse of Statistics in Social Work. KATE HOLLADAY CLAGHORN.
Strike Statistics. IRA CROSS.
Federal Census Reports: Statistics of Cities, 1905. EDWARD M. HARTWELL. $1.00.

No. 83. A Statistical Study of Infant Mortality. EDWARD BUNNELL PHELPS.
The Statistical Study of Causes of Destitution. GUSTAVE KLEENE. 75 cents.

No. 84. Uniformity and Co-operation in the Census Methods of the Republics of the American Continent. S. N. D. NORTH.
Occupation Mortality Statistics of Sheffield, England, 1890-1907. F. S. CRUM.
Standardization of Housing Investigations. JOHN R. COMMONS.
Devices for Avoiding Error. GEORGE K. HOLMES.
The Cost of Municipal Government in Massachusetts. EDWARD M. HARTWELL. 75 cents.

Digitized by Google

No. 85. State Pensions and Annuities in Old Age. FREDERICK L. HOFFMAN.
Massachusetts Savings-Bank Insurance and Pension System. LOUIS D. BRANDEIS.
The Work of the Massachusetts Commission on Old Age Pensions. F. SPENCER BALDWIN.
The Relation of Statistics to Economics and Sociology. S. N. D. NORTH. $1.00.

No. 86. The Life and Work of Carroll Davidson Wright. S. N. D. NORTH.
The Federal Census of Occupations. WILLIAM C. HUNT.
Statistics of Divorce. JOSEPH A. HILL.
Better Statistics of Industrial Mortality for the United States. CRESSY L. WILBUR.
The Teaching of Statistics. WILLIAM B. BAILEY. $1.00.

No. 87. The Extent of Unemployment in the United States. SCOTT NEARING.
California Vital Statistics. ALLYN A. YOUNG.
Bibliography of the Writings of Hon. Carroll D. Wright, 1874-1908. 50 cents.

No. 88. Industrial Accidents and Industrial Diseases. FREDERICK L. HOFFMAN.
Proceedings of the Quarterly Meeting of the American Statistical Association, Washington, D. C., September 24, 1909.
Introductory Remarks. HARRY T. NEWCOMB.
Census Methods. HON. E. DANA DURAND.
A Statistical Pilgrimage. CRESSY L. WILBUR.
A Study of New England Mortality. HARRY A. RICHARDS. $1.00.

VOL. XII. 1910-1911. Nos. 89-96. Pp. 854. PRICE $4.00.

No. 89. The Findings of the Massachusetts Commission on Old Age Pensions. F. SPENCER BALDWIN.
Professor Fisher's Formula for Estimating the Velocity of the Circulation of Money. DAVID KINLIY.
Proceedings of the Seventy-First Annual Meeting of the American Statistical Association, New York, December 27-30, 1909.
The Outlook for American Statistics. WALTER F. WILLCOX.
Changes in Census Methods for the Census of 1910. HON. E. DANA DURAND.
The Statistical Basis of Budget-Making. HERMAN A. METZ. $1.00.

No. 90. Derivation of the United States Mortality Table by Osculatory Interpolation. JAMES W. GLOVER.
The Age Returns of the Twelfth Census, 1909. WILLIAM B. BAILEY AND JULIUS H. PARMELEE.
The Social Marking System. FRANKLIN H. GIDDINGS.
The Necessity of the Supervision of Weights and Measures. FRITZ REICHMANN.
Organized Labor and the Recent Advance in Prices. LEWIS H. HANEY.
A Formula for Drawing Two Correlated Curves so as to Make the Resemblance as Close as Possible. FRANCIS TODD H'DOUBLER. $1.00.

No. 91. International Crop-Reporting Service. C. C. CLARK.
Railway Statistics. LEWIS H. HANEY.
Referenda in Massachusetts and Boston. EDWARD M. HARTWELL. $1.00.

Digitized by Google

No. 92. The Correlation of Economic Statistics. WARREN M. PERSONS.
Scope and Methods of Presentation of the Results of the Thirteenth Census of Population. W. F. WILLOUGHBY.
A Statistical Survey of Infant Mortality's Urgent Call for Action. EDWARD BUNNELL PHELPS.
The Census Age Question. ALLYN A. YOUNG.
The New York Budget Exhibit. LEONARD P. AYRES. $1.00.

No. 93. Vermont. An Historical and Statistical Study of the Progress of the State. BY WILLIAM S. ROSSITER.
Proceedings of the Annual Meeting of the American Statistical Association, 1910.
Degree of Accuracy in Census Statistics of Agriculture. BY L. G POWERS.
The Census as a Source Book in Agricultural Economics. BY H. C. TAYLOR.
Farm Tenancy in Iowa. BY BENJAMIN H. HIBBARD.
Changes in Land Values, Farms, Tenants and Owners Since 1900. BY J. L. COULTER.
Changes in Agricultural Conditions. BY J. G. THOMPSON.
Changes in Minnesota Agriculture. BY E. V. D. ROBINSON.
Large-Scale and Small-Scale Farming. BY T. N. CARVER.
The Change in the Proportion of Children in the United States and in the Birth Rate in France During the Nineteenth Century. BY W. F. WILLCOX.
The Census Age Question: A Reply. BY W. B. BAILEY AND J. H. PARMELEE. $1.00.

No. 94. Vital and Monetary Losses in the United States Due to Typhoid Fever. BY WILLIAM O. MENDENHALL AND EARL W. CASTLE.
Movement of Wholesale Prices in New York City, 1825-1863. BY CARL H. JUERGENS.
Medical and Physical Examination of School Children. BY FREDERICK L. HOFFMAN
Some Recent Developments of School Statistics. BY ROLAND P. FALKNER.
Woman and Child Workers in Cotton Mills. BY WALTER B. PALMER.
Classification of Occupations. BY ALBA M. EDWARDS.
Industrial Accidents, Employer's Liability, and Workmen's Compensation in Minnesota. BY DON D. LESCOHIER. 75 cents.

No. 95. Fifty Years of American Life Insurance Progress. BY FREDERICK L. HOFFMAN.
The Share of Vermont in the Production of Distinguished Men. BY FREDERICK ADAMS WOODS. $1.00.

No. 96. Marital and Occupational Statistics of Graduates of Mount Holyoke College. BY AMY HEWES.
Mental Ability in Relation to Head Circumference, Cephalic Index, Sociological Condition, Sex, Age, and Nationality. BY ARTHUR MACDONALD.
Two Methods of Estimating the Growth of Criminality in the United States. BY LOUIS NEWTON ROBINSON.
Fisher's The Purchasing Power of Money. BY WARREN M. PERSONS. 50 cents.

VOL. XIII. 1912-1913. Nos. 97-104. Pp. 657. Price $4.00.

No. 97. The Increase in Industrial Accidents. BY HENRY J. HARRIS, PH.D.
Proceedings of the Seventy-Third Annual Meeting of the American Statistical Association, Washington, D. C., December 28-29, 1911.
Ascertaining and Forecasting Business Conditions by the Study of Statistics. BY ROGER W. BABSON.

Digitized by Google

The Rural South. By JOHN LEE COULTER, PH.D.
Southern Agriculture and the Negro Farmer. By H. B. FRISSELL, LL.D.
Rural Life in the South. By E. C. BRANSON, A.M.
Rural Conditions in the South. By WILLIAM H. GLASSON.
The Rural South. By W.E. DUBOIS, PH.D.
The Decline in Rural Population. By BENJAMIN H. HIBBARD, PH.D.
The Decline of the Rural Population in New England. By ALEXANDER E. CANCE, PH.D. 75 cents.

No. 98. Reporting of Industrial Accidents. By ROBERT EMMET CHADDOCK, PH.D.
The Coal Supplies and Coal Production of the United States. By EDWARD W. PARKER.
Wages in Massachusetts and New Jersey. By SCOTT NEARING, PH.D.
The Degree of Correspondence Between Two Series of Index Numbers. By J. D. MAGEE, A.M. 75 cents.

No. 99. Statistics at the Fourteenth International Congress on Hygiene and Demography, Berlin, September 23-29, 1907. By WALTER F. WILLCOX.
The Permanent Census Board of New York City: Its Present Work and Possibilities for Development. By GEORGE H. CHATFIELD.
The Mortality of the Workmen's Sick and Death Benefit Fund of America. By EDWARD H. KONIGER. 75 cents.

No. 100. The Use of Averages in Expressing the Wages and Hours of Milwaukee Street Car Trainmen. By HORACE SECRIST.
Statistics at the Fifteenth International Congress on Hygiene and Demography, Held at Washington, September 23-28, 1912. By WALTER F. WILLCOX. 50 cents.

No. 101. The Need of Social Statistics as an Aid to the Courts. By WALTER F. WILLCOX.
Proceedings of the Seventy-fourth Annual Meeting of the American Statistical Association, Boston, Mass., December 27-31, 1912.
The Function of the State in Relation to Statistics of Municipal Finances. By CHARLES F. GETTEMY.
Unit Accounting in Social Work. By ROBERT A. WOODS.
Some Possibilities in the Practical Application of Federal Census Results. By WILLIAM S. ROSSITER.
Some Recent Changes in the Composition of the Population of the United States. By WILLIAM B. BAILEY.
The German Statistical Society and its Annual Meeting in Berlin, 1912. By EUGEN WÜRZBERGER.
A Measure of the Manner of Living. By C. A. PERRY. 75 cents.

No. 102. International Statistics of the Consumption of Alcohol. By DR. E. W. MILLIET.
Benevolent Hospitals in Metropolitan Boston. By WILLIAM H. MAHONEY.
Street Accidents—New York City. By E. S. CLOWES.
Occupational Mortality and Causes of Death. By DR. JACQUES BERTILLON.
"Gresham's Law" as Applied to Immigration to Hawaii. By VICTOR S. CLARK. 50 cents.

No. 103. Street Traffic Accidents. By FREDERICK S. CRUM.
The International Statistical Institute, XIVth Session, Vienna, September, 1913. By JOHN KOREN.
An Index of Changes in Extractive Industries. By WILLIAM E. LEONARD. $1.00.

Digitized by Google

No. 104. Our Coming Seventy-fifth Anniversary. By JOHN KOREN.
The Census Methods of the Future. By E. DANA DURAND.
Comparative Fecundity of Women of Native and Foreign Parentage in the United States. By JOSEPH A. HILL.
The Permanent Census Bureau: A Decade of Work. By JOHN CUMMINGS.
An Experience in the Compilation of Mortality Statistics. By LOUIS I. DUBLIN AND EDWIN W. KOPF.
The Mother Tongue Inquiry in the Census of Population. By E. A. GOLDENWEISER. $1.00.

VOL. XIV. 1914-1915. Nos. 105-112. Pp. 831. PRICE $4.00.

No. 105. The Seventy-fifth Anniversary Celebration of the American Statistical Association.
The Service of Statistics to Economics. By DAVID KINLEY.
The Service of Statistics to Sociology. By FRANKLIN H. GIDDINGS.
The Service of Statistics to History. By CHARLES H. HULL.
The Service and Importance of Statistics to Biology. By RAYMOND PEARL.
The Technique of Public Statistical Exhibits. By CHARLES J. STOREY.
The Influence of Marriage on the Death-rate of Men and Women. By GEORGE I. BLISS.
The Need of a Federal Census. By MELVIN T. COPELAND.
The Number of Jews in New York City. By HENRY CHALMERS. $1.00.

No. 106. The Present Status of Statistical Work and How It Needs to be Developed in the Service of the Federal Government. By WILLIAM S. ROSSITER.
Present Status of Statistical Work and How It Needs to be Developed in the Service of the States. By ADNA F. WEBER.
Statistics in the Service of the Municipality. By F. SPENCER BALDWIN.
The Present Status of Statistical Work and How It Needs to be Developed in the Service of Private Societies and Organizations. By W. S. GIFFORD.
A National Budget. By HARVEY S. CHASE.
Methods of Direct Legislation in Oregon. By WILLIAM F. OGBURN.
Education and Fecundity. By NELLIE SEEDS NEARING. 75 cents.

No. 107. The Occupation Hazard of Locomotive Firemen. By HENRY J. HARRIS.
Vital Statistics—The White Slave of Sanitation. By CRESSY L. WILBUR, M.D.
The Decadence of the Native American Stock. A Statistical Study of Genealogical Records. By FREDERICK S. CRUM.
Immigration as a Source of Urban Increase. By F. STUART CHAPIN, PH.D.
American Life Tables. By C. H. FORSYTH.
Service Income and Property Income. By SCOTT NEARING. 75 cents.

No. 108. Coöperation Between Academic and Official Statisticians. By WALTER F. WILLCOX.
The Economic Progress of the United States During the Last Seventy-five Years. By FREDERICK L. HOFFMAN.
Records of Health and Sanitary Progress. By ROBERT E. CHADDOCK.
Some Census Publications and Census Methods. By EDWARD M. HARTWELL. 50 cents.

Digitized by Google

No. 109. Some Statistical Ideals. By JOHN KOREN.
A Standard Accident Table as a Basis for Compensation Rates. By I. M. RUBINOW.
The Statistical Work of the United States Government. By WALTER F. WILLCOX.
How the Statistical Output of Federal Bureaus Might be Improved. By W. C. MITCHELL.
The Statistical Work of the United States Government. By E. DANA DURAND.
Some Features of the Statistical Work of the Bureau of Labor Statistics. By ROYAL MEEKER.
Coöperation of Federal Bureaus with Private Agencies in Statistical Work. By JOHN CUMMINGS.
Some Present Statistical Needs and the Statistical Work of the Federal Government. By W. S. GIFFORD.
Concerning Uniform International Financial Statements. By HARVEY S. CHASE.
International Coöperation for the Standardization of Statistical Work. By ROGER W. BABSON.
The Census Office in Commission. By S. N. D. NORTH.
Proceedings of the Seventy-Sixth Annual Meeting of the American Statistical Association, Princeton, N. J., December 28-31, 1914.
$1.00.

No. 110. Public Service Statistics in the United States. By JULIUS H. PARMELEE.
Infant Mortality in Fall River, Massachusetts—A Survey of the Mortality among 833 Infants born in June, July, and August, 1913. By LOUIS I. DUBLIN.
Income Tax Statistics. By ROLAND P. FALKNER.
Old Age and the Industrial Scrap-Heap. By ARTHUR J. TODD.
Estimates of a Living Wage for Female Workers. By CHARLES E. PERSONS.
The Improvement and Extension of the Registration Area. By LOUIS I. DUBLIN.
Osculatory Interpolation Formulas. By C. H. FORSYTH.
$1.00.

No. 111. The Social Survey and Its Further Development. By J. L. GILLIN.
Data on Unemployment from Employers' Records. By HOWARD WOOLSTON.
The Value to Economics of Formal Statistical Methods. By CARL J. WEST.
Infant Mortality and the Size of the Family. By HENRY H. HIBBS.
Measure of Rural Migration and Other Factors of Urban Increase in the United States. By JOHN M. GILLETTE AND GEORGE R. DAVIES.
Contributions to Urban Growth. By EARLE CLARKE.
A Study of the Causes of Industrial Accidents. By GUSTAVUS MYERS.
75 cents.

No. 112. The Service of Statistics in Problems of War and Peace. By E. DANA DURAND.
Some Population Statistics of the Pacific Coast States. By WALTER F. WILLCOX.
"On the Handicapping of the First Born," a Criticism of Professor Pearson's 1914 Memoir. By LOUIS I. DUBLIN AND HARRY LANGMAN.
Vital Statistics Work in California. By GEORGE D. LESLIE.
Theory of Statistical Tabulation. By G. P. WATKINS.
Vital and Monetary Losses in the United States Due to Preventable Deaths. By C. H. FORSYTH.

Digitized by Google

Joint Committee on Standards for Graphic Presentation.
New Method for Computing the Moving Average. By Willford I. King.
Comparative Militarism. By Arthur MacDonald.
Wholesale Prices for the United States, 1801-1840. By Alvin H. Hansen.
The Present Position of Infant Mortality: Its Recent Decline in the United States. By Henry Horace Hibbs, Jr. $1.00.

VOL. XV, 1916–1917, not yet completed.

No. 113. The Nature and Significance of the Changes in the Birth and Death Rates in Recent Years. By Walter F. Willcox.
Statistics of Imports and Exports. By Frank R. Rutter.
The Mortality of Our Public Men. By Irving Fisher.
The Agricultural Element in the Population. By Eugene Merritt.
The Mother and Infant Mortality. By Henry H. Hibbs, Jr.
Measures of Relative Tax Burdens. By A. E. James.
Statistics of the Food Supply in Germany. By Robert Morse Woodbury.
Proceedings of the Seventy-Seventh Annual Meeting of the American Statistical Association, Washington, D. C.; December 28-30. 1915. 75 cents.

No. 114. Some Problems of Population Growth. By E. Dana Durand.
The Budgets of Smith College Girls. By F. Stuart Chapin.
A Comparison of the Relative Healthfulness of Certain Cities in the United States Based upon the Study of Their Vital Statistics. By G. E. Harmon.
The Improvement of Statistics of Cause of Death Through Supplementary Inquiries to Physicians. By Louis I. Dublin and Edwin W. Kopf.
Statistical Tabulation and Practice. By Roland P. Falkner.
Statistical Standardization in Washington. By Julius H. Parmelee.
Classification and Limitations of Statistical Graphics. By E. A. Goldenweiser.
The Use of the Correspondence Method in Original Research. By Asher Hobson. 75 cents.

No. 115. Scope and Method of Statistics. (Introduction by Walter F. Willcox.) By Harald Westergaard.
Comment on Westergaard's "Scope and Method of Statistics." By Warren M. Persons.
Comment on Westergaard's "Scope and Method of Statistics." By Julius H. Parmelee.
Comment on Westergaard's "Scope and Method of Statistics." By Carl J. West.
Comment on Westergaard's "Scope and Method of Statistics." By G. P. Watkins.
The Federal Farm Loan Act. By Walter B. Palmer.
Wages in American Cities. By W. F. Ogburn and Esther Kelly.
Highway Accidents in New York City During 1915. By Mills E. Case.
Regularity of Employment—Need of Standardization of Methods for Measuring Regularity of Employment. By N. I. Stone. $1.00.

No. 116. A Study in Social Dynamics: A Statistical Determination of the Rate of Natural Increase, and of the Factors Accounting for the Increase in Population in the United States. By John M. Gillette.
Forty-five Years of U. S. Naval Morbidity Statistics of Degenerative Disorders. By Charles N. Fiske.
Florence Nightingale as Statistician. By Edwin W. Kopf.
Consumption of Gas as Affected by Population. By G. P. Watkins and L. H. Lubarsky.

Digitized by Google

Interpolation of Ordinates and Areas among Areas. By C. H. Forsyth.

Charters for American Business Corporations in the Eighteenth [Century. By Joseph S. Davis.

The Application of Statistics to Advertising and Marketing. By Hyman L. Roth. $1.00.

Address CARROLL W. DOTEN, Secretary,
491 Boylston Street, Boston, Mass.

Digitized by Google

Digitized by Google

AMERICAN
STATISTICAL ASSOCIATION

NEW SERIES, No. 117. JUNE, 1917.

THE UTILIZATION OF STATISTICS IN BUSINESS.*

By Julius H. Parmelee, Ph.D., *Statistician Bureau of Railway Economics.*

INTRODUCTORY.

Let me preface this paper with a brief quotation from Westergaard's treatise on "Scope and Method of Statistics," which appeared in the Quarterly Publications of this Association last June. "A prominent feature of modern statistics is the immense development of the field. . . . In the future the difficulty will be not so much in gathering material as in mastering it, in digesting all these masses of reports which have been stored in the archives and on the book-shelves of statistical offices."

In the future the difficulty will be not so much in *gathering* material as in *mastering* and *digesting* it. These words strike a definite keynote in regard to the modern problem of statistics. A good digestive apparatus thoroughly utilizes the material offered to it, discarding what is unnecessary or ill-adapted and reconstructing what is useful into the form best suited to the general organism of which it is a function. Similarly, the statistician of today must utilize the material at his disposal first by discarding the useless, the irrelevant, and the defective, and then by rearranging the remainder into the form and substance best suited to the needs of the organization with which his interests are bound up.

Statistics are now collected and compiled on every conceivable subject under the broad heavens. This activity is carried

* Paper presented at the annual meeting of the American Statistical Association, Columbus, Ohio December 28, 1916.

Digitized by Google

on in large part by means of governmental, but in part also by private or voluntary organizations. The principal problem of today is to utilize the great mass of material already available, together with what is constantly being added, so as to make it of the greatest service to society.

Restricting ourselves to the question of statistics as utilized in the modern business world, we may ask: First, what organizations collect statistics of interest to business; Second, what use is actually made of statistics in business practice.

By "business" I mean in a general way all forms of effort put forth by modern society primarily and directly for economic gain. This definition excludes such activities as are devoted to general informational, political, educational, and scientific purposes, even though many of these activities may be, and some undoubtedly are, of indirect service to business. In other words, I shall confine myself strictly to the direct and conscious efforts of the business world to secure and utilize statistical information for its own economic and financial benefit. Even thus restricted, the field presents a much greater extent than could possibly be covered within the limits of a brief paper, and I shall attempt little more than to indicate the several lines along which development seems to be tending most rapidly today.

ORGANIZATIONS THAT COLLECT BUSINESS STATISTICS.

It is clear that the primary sources of information of service to the business world are those afforded by governmental agencies; in especial, the agencies of the federal government. Without going into detail, it may be said that the Departments of Commerce, Agriculture, and the Treasury, and the Interstate Commerce Commission, are the principal federal agencies compiling business statistics, although we must not overlook the Department of Labor, various independent boards such as the Federal Reserve Board, Federal Trade Commission, Farm Loan Board, the newly created shipping and tariff boards, and the like. My effort here, however, is to indicate the compilation activities of private rather than of governmental or public agencies. The number of private agencies is legion, and can be discussed briefly only by general grouping.

Digitized by Google

First, we have organizations maintained individually or coöperatively by industrial corporations, whose investigations and compilations are designed to cover the whole of the industry in which the particular corporations are engaged. Perhaps the oldest of such organizations now in existence is the American Iron and Steel Institute, which was established in 1855, under a slightly different name, and has published statistical reports of the iron and steel industry ever since. One clause of the constitution adopted in 1855 read as follows: "The general objects of this association shall be to procure regularly the statistics of the trade both at home and abroad." This clause was retained in the constitution when the American Iron and Steel Institute replaced the older association, and may be found there today. In pursuance of this constitutional provision for the collection of statistics of the industry, the secretary of the association compiled in 1859 a guide to the "iron works and iron ore mines of the United States." This was the only available compilation of data regarding American furnaces, rolling mills, forges, and steel works until 1879, when the association published the first edition of its directory to the iron and steel works of the United States, which has been reissued from time to time, down to the present year. The American Iron and Steel Institute in 1913 established a statistical bureau, which issues annual statistical reports covering the production, imports and exports, and prices of iron and steel, iron ore, coal and coke. In addition, the bureau issues bulletins from time to time giving current statistics of production of pig iron, steel ingots and castings, steel rails, nails, pipe, plates, etc.

I have described the statistical work of the Iron and Steel Institute in some detail, both because it is probably the oldest existing organization of the kind, and because it covers the field represented by its industry in an unusually complete and satisfactory manner. Coöperative statistical bureaus or organizations of some kind are maintained in a number of other industries, besides that of iron and steel manufacture. The form, organization, and title of these bureaus vary considerably from industry to industry, but they have this one feature in common: they collect and compile statistics dealing

Digitized by Google

with their respective industries. Such bureaus or offices exist in the electric railway industry; among life insurance corporations; and in the steam railway industry. In addition, a number of national associations representing particular industries have among their aims the collection and dissemination of statistical data regarding their respective activities. In especial, I would mention the American Forestry Association, which compiles forest statistics; the United States Brewers' Association, which publishes a yearly handbook on the status of the trade; the American Foundrymen's Association, organized in part to collect all proper information connected with the foundry business; the American Mining Congress, which strives to disseminate information in relation to mining, metallurgy, and allied industries; the Copper Producers' Association, which compiles copper statistics; the American Water Works Association, the National Association of Cotton Manufacturers' which collects and imparts information relating to the cotton industry; the Street Railway Association of the State of New York, organized for the acquisition and dissemination of experimental, statistical, and scientific knowledge relating to the construction, equipment, and operation of street railways.

Among the statistical offices created by and among the steam railways are two committees of the American Railway Association, one on relations between railroads and the other on accident statistics. The first of these committees was created in 1907 and has been publishing statistics of freight car surpluses and shortages ever since that date; also, a part of the time, statistics of freight car performance. The other committee was created in 1915 for the purpose of making suggestions for the improvement of the methods of reporting and compiling statistics of railway accidents. Other statistical bureaus organized by the railways have been the Special Committee on the Relation of Railway Operations to Legislation, which has been in existence since 1909 and publishes statistics of laws introduced and passed by state legislatures each year relating to railway operations; the Bureau of Railway News and Statistics, maintained in Chicago by a group of railways, which has been in existence since 1904 and which

Digitized by Google

compiles annually and periodically statistics of various aspects of the railway question; the Bureau of Railway Economics, organized at Washington in 1910 by railways of the United States, whose work is largely statistical, consisting of various publications on railway revenues and expenses, traffic, stockholders, comparative railway statistics of the United States and foreign countries, wages, equipment, etc. This bureau also compiles statistical exhibits for rate cases, wage arbitrations, and the like, and prepares statistical material on various topics as the result of special inquiries.

A number of large coöperative bodies of business men also maintain organizations for the compilation and collection, or in some cases the improvement, of business statistics. In this class is the Committee on Statistics and Standards of the Chamber of Commerce of the United States recently created. This national federation of local bodies of business men—the United States Chamber of Commerce—was organized in April, 1912. Among other functions it utilizes the commercial data gathered by government bureaus by directing it into the channels to which it is immediately applicable, and analyzes statistics with regard to the production and distribution of manufactures at home and abroad. In other words, the chamber digests and adapts existing business information to the needs of the business bodies constituent to it.

In the second place, we have statistical organizations established and maintained by individual corporations. There are so many of these that I can allude only briefly to a few, which may be regarded as representative of the general situation. First of all may be mentioned the statistical organizations of various brokerage and financial firms, including the larger banks of the country. An excellent example of statistical organizations maintained by banks is that of the National City Bank of New York. The statistical work of that bank consists primarily of the preparation of statistical statements in response to calls from officers of the bank regarding commercial, financial, and industrial conditions in the United States and all other sections of the world. In addition, there is prepared each week a statistical analysis

Digitized by Google

of the import and export trade of the port of New York. While these statements are prepared primarily for the officers of the bank, they are also supplied freely, in the interest of foreign commerce, to any one making application for them.

Other large banks have statistical departments, as do nearly all of the important banking and brokerage houses in the large cities. Speaking generally, these statistical departments keep track of occurrences and possibilities in the industrial and financial field, and present their findings in general or special reports.

Many manufacturing and industrial firms have departments that compile statistical information along their particular lines. Among organizations which have done this are banks; brokerage and banking houses; local boards of trade, chambers of commerce, stock and produce exchanges, etc.; life and fire insurance companies; telegraph and telephone companies; makers of tires and tire fabrics; locomotive manufacturers; mercantile establishments purveying various lines of goods; construction companies; makers of bank note paper; hardware and brass manufacturers; gas, electric light, heat, and power companies; cotton and silk manufacturers; street and interurban railways; pottery makers; manufacturers of farm implements; tobacco manufacturers; makers of storage batteries and general electric equipment; research laboratories; manufacturers of soda ash; munitions makers; paper pulp makers; taxicab companies. This is only a partial list, chosen more or less at random; yet the diversity of interests represented in the list clearly indicates the extent to which the modern business world is utilizing statistics.

Prominent among individual corporations that have established statistical organizations are the steam railways of the United States. Hardly one of the larger railways but maintains a statistical office of some kind, either in connection with its accounting department or as an entirely separate branch. These statistical offices usually prepare the annual and other periodical reports required by various govermental agencies; many of them compile much additional information for the executives in charge of their operations. Some of the large roads carry this statistical work to a very high

Digitized by Google

degree of detail and completeness. From these the list shades down to roads which have no statistical force at all, so called, yet are compelled by force of law or circumstances to do a certain amount of statistical work.

A third type of private organization engaged in compiling business statistics consists of the group established primarily for the purpose of collecting and distributing, in digested form, the fundamental statistics of the business or financial world. I refer to the statistical or ecomonic services that prepare statistical charts, barometric diagrams, and the like, for the information of their clients, who are usually business men with a keen interest in industrial possibilities. Business men who subscribe to these services are often of the class who can not afford to maintain a separate statistical organization, or even to keep a statistician on whole or part time, but who can afford to pay a reasonable annual fee for such material as is prepared by the services.

One organization of this type has been doing business since 1902, and another since 1906, although the larger development of this kind of service has been comparatively recent. If I understand the underlying principle of this class of services aright, it is that business conditions move in cycles; that these cycles are not of uniform length or duration; yet that there are certain positive signs of the progressive development of each cycle, which can be observed only by means of a careful charting of the business and financial situation; that by accurate reading of these signs warning can be given of coming events—whether good or bad.

Perhaps a brief description of the kind of work done by organizations of this kind will give the clearest impression of their nature, and the significance of their statistical work. One service publishes periodical reports designed to "give merchants, bankers, and investors carefully collected and condensed facts about business and finance." Among these reports are a so-called "weekly barometer letter," analyzing and forecasting fundamental probabilities in business, and indicating the position of business and finance in the current cycle; a monthly desk sheet, giving basic statistics as to new building, bank clearings, railway operations, crops, etc., de-

Digitized by Google

signed to indicate the fundamental condition of American business; a periodical report on mercantile and commodity price conditions.

Another service, using practically its own words to describe its activities, brings together, in a barometer chart, composite cycles indicating the actual condition of business and finance and forming a basis for forecasting any impending change; serves as a clearing house for a great mass of statistical data, some of which is fundamental and some only of record significance, but all of which can be and is applied to solve the problems of business men, bankers, private investors, and the like; compiles and analyzes special statistics for separate industries.

Still another service is called an "investors' service," which supplies to clients a number of periodical statistical statements on financial conditions, investments, general business, railways and industrial corporations, and the like. This service also issues a monthly business barometer, based on the belief that "all things move more or less in cycles, and to understand these cycles involves a continuous study and analysis of fundamental business conditions."

I have devoted what may seem like a disproportionate amount of attention to the work of these statistical services or clearing houses, not because I am in any way interested in their welfare, but because they seem to be making an effort, at least, to live up to the Westergaard motto with which we opened this paper. In other words, they are taking a great mass of statistical material already available, gathered by other persons or organizations, and are endeavoring to master and digest it, so as to be of the greatest service to the greatest number. The work of all organizations of this sort is still necessarily somewhat in the pioneer stage. Part of their future development will doubtless come from the internal growth and improvement of the organizations themselves, and part from a greater willingness of industrial concerns to furnish figures long regarded as trade secrets. However it may come, development of work of this kind appeals to me as being distinctly along the line of making business statistics really serviceable to society.

Digitized by Google

In addition to services of the kind I have described, there are many organizations and individuals engaged in the same sort of work, although on a smaller scale. Some men hold themselves out as expert compilers or statisticians, consulting statistical experts, and the like. In much of this work there is unquestionably room for improvement; but there is definite promise for the future in the development of all organizations designed to make statistics serviceable to the business world.

UTILIZATION OF STATISTICS.

We come now to the second branch of our inquiry. The first dealt with the growth of organizations for the compilation of business statistics. The second will be devoted to the use actually made of the statistics when compiled.

It is clear, almost without reflection, that a business man or corporation will use statistics to guide him in his business decisions; to tell him when to expand his activities and when to contract them; to inform him in what sections to look for the safest and quickest results from a sales campaign; to help him decide as to the advisability of a particular proposition; in fact, to guide him daily in the various decisions that are constantly calling for attention.

Further reflection will make the possibilities along these lines stand out even more definitely. Suppose we illustrate these possibilities by a few concrete examples from the business world of today. Take for example, corporations maintaining a national chain of stores, such as the United Cigar Stores, the F. W. Woolworth Company, or any other of similar nature. It is clear that these corporations must keep in close touch with statistics of supply and especially of demand, both by means of reports secured direct from their constituent stores and by observation of general market and purchasing conditions. Going a step further, the newly organized American International Corporation must now have, or must certainly establish, a statistical organization covering not only the situation in the United States but also the whole civilized world as well. Again, every bank and banking house must keep in touch with business and financial conditions and possibilities by statistical means. Whether banks maintain a statistical

Digitized by Google

department or not, they are increasingly forced to keep track of the future by means of statistics.

Similarly, all business and mercantile houses keep an eye to the future by watching the business barometer for their particular line of activity. A certain hardware manufacturing corporation of the middle west, for example, compiles statistics of crop and farm conditions, as a clue to the purchasing power of the agricultural districts. Trade statistics, banking and financial statistics, agricultural statistics, and the like are all a part of the equipment of modern business houses.

Perhaps the greatest development of the utilization of business statistics on a large scale is in the railway industry. One reason for this is the immensity of the industry, and the complexity of its functions. Another is the character of the railway organization, the necessity of its activities being spread over a wide area, and the difficulty of keeping in touch with all parts of the system without detailed statistical reports. Although the manager of a large industrial plant keeps in daily personal contact with his departmental and division chiefs, a railway general manager has his territory spread over thousands of square miles of area, and can keep in touch with his plant only by a system of daily reports. Hence arises the detailed reporting methods in vogue on all railway systems, whereby the executives receive daily statistical reports from their subordinates, covering all features of current operations.

A friend of mine tells with amusement of his conversation with a railway official. The official, in describing the work done by the various men in his organization, said the division superintendent's duties consisted in reporting to the general superintendent, the duty of the general superintendent was to report to the general manager, the general manager reported to the vice-president in charge of operations, and the vice-president reported to the president. "And to whom does the president report," asked my friend. "He reports to the board of directors," was the reply. On the face of it, these railway officials expended the whole of their energies merely preparing and presenting reports to their chiefs next in command. As a matter of fact, the incident has significance in showing how thoroughly the system of accurate records is

Digitized by Google

woven into the warp and woof of the railway business, and to a considerable extent of other industrial activity. No official could make an intelligent report to his superior were he not himself conversant with every phase of his work; and if his own work is going wrong at any point, he is not likely to turn in a report until the matter is rectified. In other words, the necessity of reporting daily the status of your work keeps you alert at all times regarding its progress and its effectiveness

SUMMARY

We have seen to how great an extent has developed the practice, on the part of industrial and financial organizations, of establishing statistical offices with the primary object of compiling business statistics of utility in their work. We have seen, further, that these statistics are put to use in numerous ways, but chiefly as guideposts along the path of industrial development and progress. These guideposts may serve as warnings in time of impending stress and storm, or as incentives to development during periods of expansion and optimism. We have noted the fact that the more complex the industry, the more serviceable is a system of statistics detailing its activities. Especially is this true of the railway industry, with its operations spread out over so vast an area.

Much the greater part of this statistical development has gone on during the past ten years. No one who has been working in the field of business statistics during recent years can fail to have been impressed by the tremendous growth in the use of statistics in the industrial world. The statistical services I have described date back hardly more than ten years; they have come into being as the result of a definite need that has only recently been asserting itself.

Take one phase of statistical growth in the railway world. The use of statistics in a wage arbitration was hardly thought of ten years ago. In 1910, in a case involving the switchmen of a group of railways, the employees put on the stand an expert statistician armed with a considerable array of statistical exhibits. From that time, the use of statistics in railway wage arbitrations has grown to tremendous proportions, costing hundreds of thousands of dollars in a single arbitra-

Digitized by Google

tion, and running the whole gamut of possible subjects. I do not unqualifiedly endorse this particular form of growth in statistical activity, much of which I consider wasteful and useless, but am merely recording it here as a matter of historical development. My emphasis is on the recent and consistent growth of this kind of statistical work. Again, the use of statistics and statistical exhibits in railway rate cases has been greatly developed during the past six or eight years.

To say that industrial complexity will increase with the future development of economic activity is to point out, in another way, the growing need of the future for accurate systems of statistical records and compilations. Along what lines the development will proceed is beyond the scope of the present discussion to conjecture; suffice it that the development must be steady and effective if it is to measure up to the growing need for accurate information. And herein lies the mission of the statistician of today and the near future: So to develop the statistical methods and results of the present that they may meet the increasingly complex needs of future industry. We must especially train ourselves, as students of business problems, to master and digest the material made increasingly available to us, that we may make it of the greatest utility to the business world of the morrow.

Digitized by Google

THE "RATIO" CHART

For Plotting Statistics.

By Professor Irving Fisher, *Yale University.*

COMPARING TWO MAGNITUDES BY THEIR DIFFERENCE AND BY THEIR RATIO.

In the last few years there has been a great increase in the appreciation and use of statistical charts. As a consequence, a number of efforts have been made to improve the technique of graphic representation in order that the chart may convey its message more quickly, accurately, and unmistakably. Toward this end Mr. Willard C. Brinton in particular has put us all under obligation.*

In the present article, I shall describe the nature and advantages of one form of chart on which to plot a statistical curve,—by which is meant any curve or broken line showing the different numerical values of a statistical magnitude at different periods of time. This chart will be found of very great assistance when the only object or the chief object is to display and compare *ratios*. It may therefore be called the "ratio" chart.†

The chief reason why ratio charting has not yet been more widely used is, I believe, that its extreme simplicity is not yet realized. Those who have mentioned it in print have usually contented themselves with stating dogmatically how it is to be used without explaining the whys and wherefores, further than that its spacing is "logarithmic" like that of a logarithmic slide rule. But most persons regard logarithms and slide rules as a species of magic and fight shy of a method, the foundations of which they do not clearly understand.

*Through the publication of his "Graphic Methods for Presenting Facts," New York (The Engineering Magazine Company), 1914, 371 pp., and through the work of the Joint Committee on Standards for Graphic Representation which grew out of the publication of his book.

†I am not, as I originally supposed, the first to hit on this simple device, although it still remains almost wholly unknown and unused and has never yet, so far as I know, been adequately described. A bibliography of the very meagre literature on the subject is given at the end of this article. It is there shown that the method in essence was used as early as 1863.

Digitized by Google

The object of this article is to make the construction of the ratio chart clear to any reader and, at the same time, to point out in some detail its various uses and advantages. In doing this it is quite unnecessary to use even the concept of logarithms.

FIG. 1*d*. FUTURE HYPOTHETICAL POPULATION OF THE UNITED STATES. ORDINARY OR DIFFERENCE METHOD.

Equal vertical intervals represent equal statistical differences. A line in ascending at a uniform *ratio* is curved. Uniformity is therefore not evident to the eye.

FIG. 1*r*. THE SAME. RATIO METHOD.

The same assumption. Uniformity is here evident to the eye, being represented by a straight line.

We may compare any two magnitudes, of like kind, by means either of their difference or of their ratio. In the first kind of comparison "an inch on the end of one's nose" is exactly as much as an inch added to the height of the Washington Monument; in the second kind of comparison, on the other hand, "an inch on the end of one's nose" is an addition of about 40 per cent., or as much as 220 feet added to the height of the Monument.

The ordinary chart is adapted to *difference* comparisons rather than to *ratio* comparisons, whereas the statistician is usually concerned with ratio comparisons far more than with difference comparisons.

CONVERTING ORDINARY PLOTTING PAPER.

Fig. 1 *d* ("d" standing for "difference") shows an ordinary plotting chart; that is, a chart with equal vertical spacing, labelled for population in millions, the labelling being for equal statistical *differences*, and each interval of verti-

Digitized by Google

cal ascent representing an increase of 10 millions over the preceding.

Since the key-idea of the ratio chart is that equal vertical intervals represent equal *ratios* of increase instead of equal differences of increase, it may also be constructed from ordinary cross ruled plotting paper. This may be done in two steps, (1) labelling the existing lines and (2) interpolating new lines. The first step is to label the existing (equidistant) horizontal lines with numbers increasing in a given ratio.

Fig. 1 *r* ("r" standing for "ratio"), shows a ratio chart made from ordinary ruled paper simply by labelling the horizontal lines for *ratios*, each interval of vertical ascent now representing an increase of 10 *per cent.* over the preceding. That is, while the first chart is labelled vertically, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, etc. (each number being 10 more than the one below) the second chart is labelled vertically, 100, 110, 121, 133, 146, 161, 177, 195, etc.* (each number being 10 *per cent.* more than the one below).

The ratio method consists simply in plotting any statistical curve by using the labels† of Fig. 1 *r* instead of those of Fig. 1 *d*.

Thus, let us plot, say, the (imaginary) population of the United States in millions, beginning with 100 millions in 1910 and assuming an increase of 10 per cent. every decade. Since all increases of 10 per cent. are represented in Fig. 1 *r* by equal vertical distances, it is clear that the curve representing population, *i. e.*, connecting the series of points in Fig. 1 *r* lying diagonally on the cross lines, will be a straight line.

Thus we see the first merit of the ratio method of charting the growth of a statistical magnitude: *uniformity* in the percentage rate of growth is pictured by *straightness* in the plotted line.

On the other hand, in the ordinary or difference chart, such as that of Fig. 1 *d*, the same population growth will be represented by an "exponential curve" and its uniformity of per-

*If we carry the numbers backward, that below 100 will be 91 (for 91:100::10:11); that next below 91 will be 83; and so on, indefinitely, down to any number, however small, except zero. Evidently we can approach as near to zero as we please but zero itself recedes into a bottomless pit; there is no base or zero line.

†Of course in the ratio chart as in the difference chart the numbering may be magnified or reduced in any ratio. Thus instead of the numbers 100, 110, 121, etc., we may substitute 1.00, 1.10, 1.21 or .100, .110, .121, etc., etc.

Digitized by Google

centage rate of growth is altogether lost to the eye. In fact the uninitiated are apt to be misled and to falsely infer from such a curve that the rate of growth is increasing.

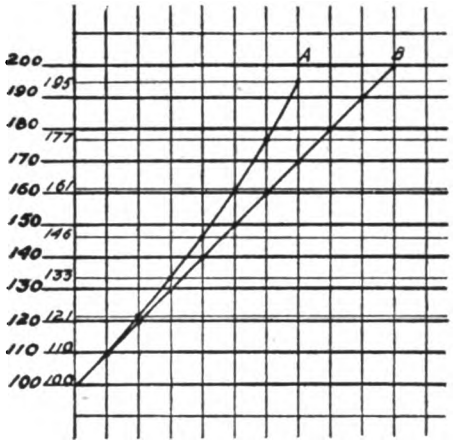


FIG. 2*d*. SHOWING GEOMETRICAL (A) AND ARITHMETICAL (B) PROGRESSIONS. DIFFERENCE METHOD.

A line (A) ascending at a uniform ratio in equal periods of time, that is, in geometrical progression is curved (upward). A line (B) ascending by equal differences in equal periods of time, that is, in arithmetical progression, is straight. Note that vertical intervals increasing in geometrical progression (see the light figures) are unequally spaced.

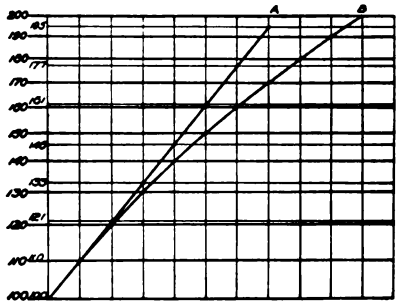


FIG. 2*r*. THE SAME. RATIO METHOD.

A line (A) ascending at a uniform ratio in equal periods of time, that is, in geometrical progression, is straight. A line (B) ascending by equal differences in equal periods of time, that is, in arithmetical progression, is curved (downward). Note that the vertical intervals increasing in geometrical progression (see the light figures) are equally spaced.

RERULING THE RATIO CHART SO OBTAINED.

But the ratio chart, as described above, is not very convenient for plotting points between the ruled lines, because so few of the labels are round numbers. To make a full fledged ratio chart, new horizontal lines need to be interpolated, at their proper places, corresponding to the round numbers 120, 130, 140, 150, etc. (Fig. 2 *r*) and the original equi-distant ones should be erased.* The contrast, then, be-

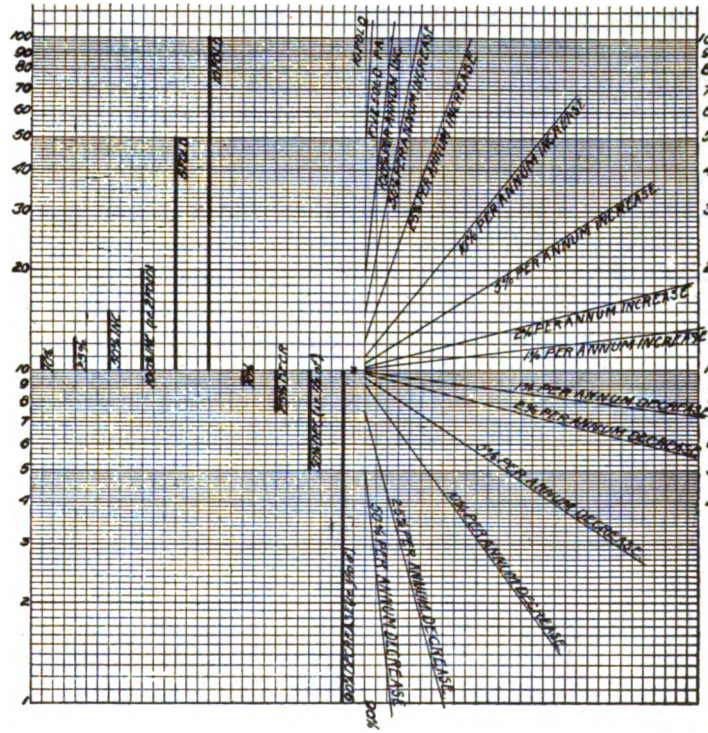
*Plotting paper for ratio charts may now be obtained ready-made from a few commercial firms, the Educational Exhibition Company, 26 Custom House Street, Providence, R. I., John Wensel, 63 West 107th Street, New York City, Keuffel and Esser Company, 127 Fulton Street, New York City, and the Standard Graph Co., 32 Union Square, New York City.

The unequally spaced horisontal lines shown in Fig. 2 *r* become more and more crowded together as we ascend, until it becomes necessary, or advisable, to omit some of them. The omission of these lines causes the plotting paper to present the appearance of an intermittent or cyclical spacing (Fig. 3 *r*).

Digitized by Google

tween the ratio chart, and the ordinary, or difference, chart, is simply one of spacing, the ratio chart (as in Fig. 2 *r*) having the numbers 120, 130, 140, etc. *unequally* spaced whereas the difference chart (as in Fig. 2 *d*) has these same numbers *equally* spaced.

The above description of a method of forming a ratio chart


100
90
80
70
60
50
40
30
20
10
9
8
7
6
5
4
3
2
1
100%
50%
25%
10%
5%
2½%
1%
½%
¼%
1/10%
1/20%
1/100%
1/200%
1/1000%
1/2000%
1/10000%
1/20000%
1/100000%
1/200000%
1/1000000%
1/2000000%
1/10000000%
1/20000000%
1/100000000%
1/200000000%
1/1000000000%
1/2000000000%
1/10000000000%
1/20000000000%
1/100000000000%
1/200000000000%
1/1000000000000%
1/2000000000000%
1/10000000000000%
1/20000000000000%
1/100000000000000%
1/200000000000000%
1/1000000000000000%
1/2000000000000000%
1/10000000000000000%
1/20000000000000000%
1/100000000000000000%
1/200000000000000000%
1/1000000000000000000%
1/2000000000000000000%
1/10000000000000000000%
1/20000000000000000000%
1/100000000000000000000%
1/200000000000000000000%
1/1000000000000000000000%
1/2000000000000000000000%
1/10000000000000000000000%
1/20000000000000000000000%
1/100000000000000000000000%
1/200000000000000000000000%
1/1000000000000000000000000%
1/2000000000000000000000000%
1/10000000000000000000000000%
1/20000000000000000000000000%
1/100000000000000000000000000%
1/200000000000000000000000000%
1/1000000000000000000000000000%
1/2000000000000000000000000000%
1/10000000000000000000000000000%
1/20000000000000000000000000000%
1/100000000000000000000000000000%
1/200000000000000000000000000000%
1/1000000000000000000000000000000%
1/2000000000000000000000000000000%
1/10000000000000000000000000000000%
1/20000000000000000000000000000000%
1/100000000000000000000000000000000%
1/200000000000000000000000000000000%
1/1000000000000000000000000000000000%
1/2000000000000000000000000000000000%
1/10000000000000000000000000000000000%
1/20000000000000000000000000000000000%
1/100000000000000000000000000000000000%
1/200000000000000000000000000000000000%
1/1000000000000000000000000000000000000%
1/2000000000000000000000000000000000000%
1/10000000000000000000000000000000000000%
1/20000000000000000000000000000000000000%
1/100000000000000000000000000000000000000%
1/200000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/200000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/1000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/2000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/10000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/20000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000%
1/100%
1/200%
1/1000%
1/2000%
1/100%
1/200%
1/1000%
1/2000%
1/1000000000000000000

represent an increase respectively of 10, 25, 50, and 100 per cent. (or 2-fold), 5-fold and 10-fold; while the next lines to the right show how far down to go to represent a decrease of 10, 25, 50 (i.e. "half of"), and 90 per cent. Any two points on the chart, however far removed from each other horizontally, if the vertical interval between them is equal to the 10 per cent. line, will be such that the statistical magnitude represented by the upper point will be 10 per cent. greater than that represented by the lower. Again the line representing "100 per cent. increase, or 2 fold" is the vertical distance between any two points on the chart of which the upper stands for twice the statistical number for which the lower stands.

The reader is advised to verify these and the other legends by taking any two points at random, reading the figures opposite them in the margin and comparing these figures.

He may also familiarize himself with the sloping lines at the right ascending respectively at the rates of 1, 2, 5, 10, 25, 50, 100 per cent. five fold, and ten fold per annum, as well as those descending at the rates of 1, 2, 5, 10, 25, 50, and 90 per cent. per annum, all drawn radiating from a center through the appropriate points on a vertical line one unit to the right of that center.

REPRESENTING GROWTH AT A CONSTANT, OR NEARLY CONSTANT, RATIO.

The advantages of the ratio chart over the difference chart are many. They may be seen from a few illustrative examples.

We have already seen (Fig. 1 *r*) that mere straightness of the plotted line indicates, in the ratio chart, a uniformity in the percentage rate of growth, whereas in the difference chart such uniformity is represented by an exponential curve (Fig. 1*d*).

In Fig. 2 *d* and Fig. 2 *r* the same lines are repeated and labelled *A*. The line *A* represents a uniform percentage rate of growth. It is an exponential curve in the difference chart (Fig. 2 *d*) and a straight line in the ratio chart (Fig. 2 *r*).

For contrast a straight line *B* is drawn on the difference chart (Fig. 2 *d*). Its straightness signifies little—merely that the same absolute difference is added each year. But the

Digitized by Google

same absolute difference is a decreasing percentage *rate* of growth and this fact is clearly interpreted in the ratio chart (Fig. 2 *r*). We may say that the chief or typical contrast between the two charts is that a straight line represents in the one an arithmetical progression and in the other a geometrical progression; which is only another way of saying that it

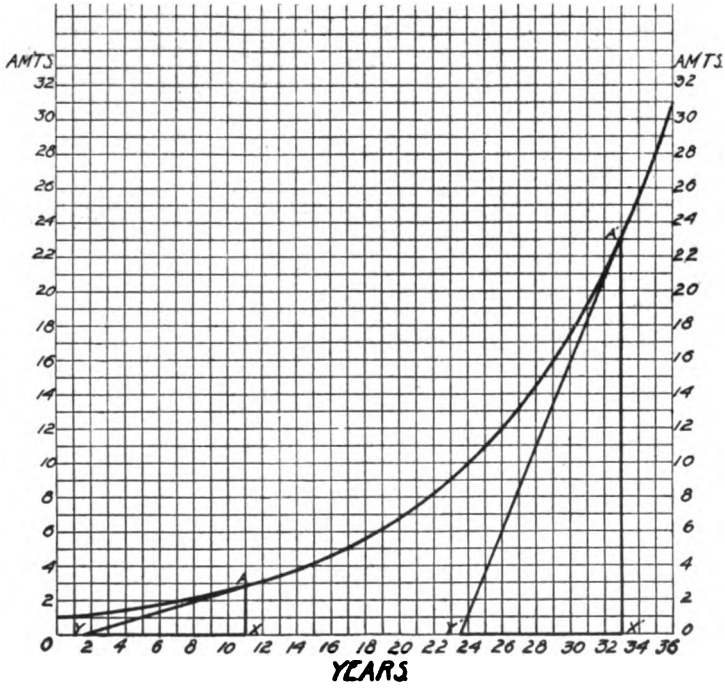


FIG. 4*d*. UNIFORM PERCENTAGE RATE. DIFFERENCE METHOD.
Showing uselessness of curve at extreme ends; also showing how to compare percentage slopes at different points by comparing (inversely) the subtangents for these points. See p. 596.

represents a progression by equal differences in the one and by equal ratios in the other.

A serious fault in the difference method is that, in a curve of rapid growth, the difference chart is useful only in the middle portion. At the extreme left such a curve, *e. g.*, an exponential, or uniform percentage, curve (as in Fig. 4 *d*) becomes almost indistinguishable from a horizontal line and at the extreme right it becomes almost indistinguishable from a vertical line.

Digitized by Google

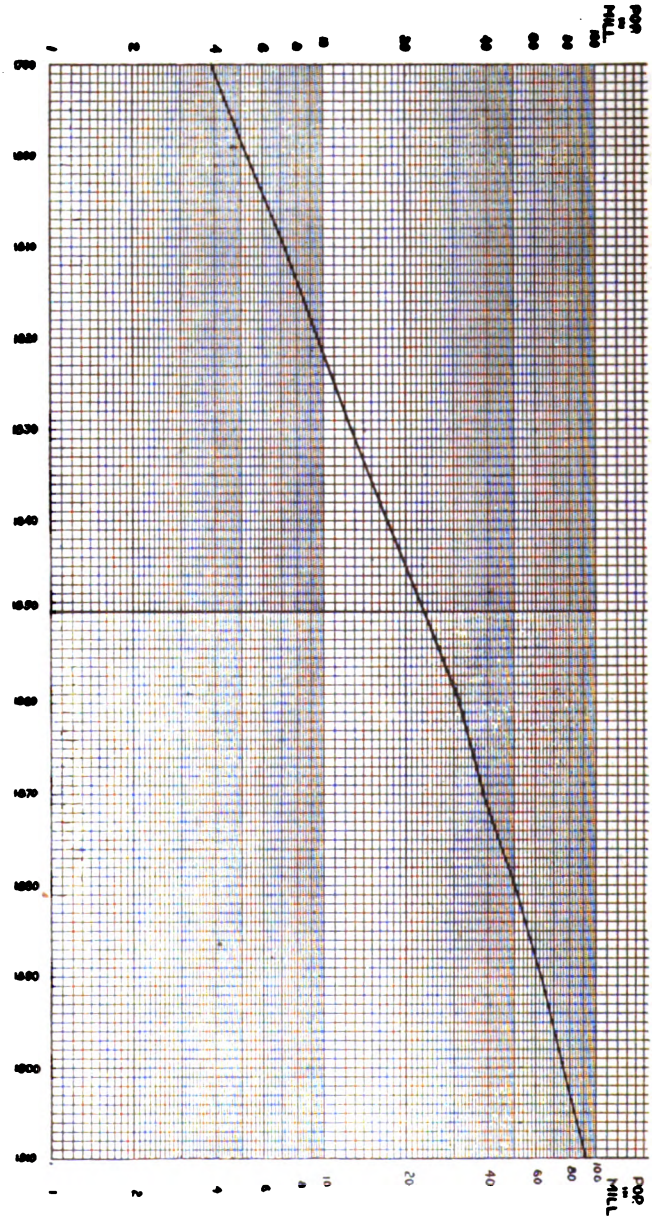


FIG. 5r. THE SAME. RATIO METHOD.
Showing clearly the slight deviations, since 1860, from a uniform rate of growth.

Digitized by Google

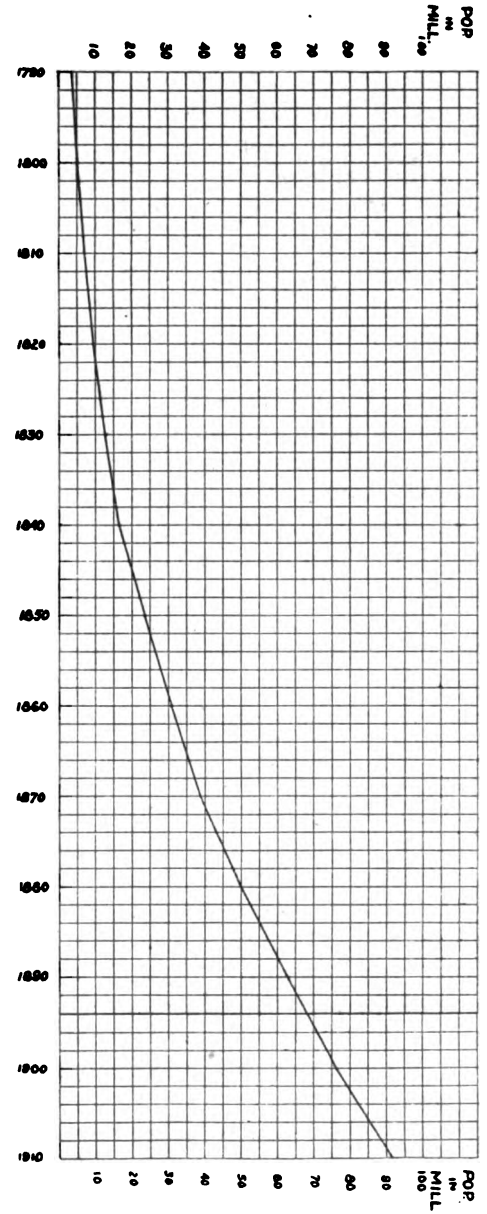


FIG. 54. ACTUAL POPULATION OF THE UNITED STATES. DIFFERENCE METHOD.
Showing the impossibility of correctly comparing rates of increase at different periods.

Digitized by Google

At either extreme no eye can estimate the percentage rate of growth although that rate may not be different from the rate in the middle.

A vast number of statistical charts represent *rapid and long continued growth*,—the statistics of a prosperous business plotted from the beginning; the statistics of a growing country; the statistics of new inventions. We have merely to mention any such familiar examples as statistics of population, wealth, crops, mining, manufacturing, railway mileage, telephones, automobiles, bank deposits, new building, sales of stocks, war debts and other magnitudes rapidly increasing since the war began, in order to realize the thousands or rather, probably, the millions of statistical charts which have been constructed of this kind, most of which are nearly useless at either end.

Thus Fig. 5 *d* represents the actual growth of the population of the United States. The reader, however discerning or experienced, cannot discover by mere ocular inspection of this curve whether the increase at the end is faster or slower than at the beginning. Fig. 5 *r*, on the other hand, shows the facts desired at a glance. It shows a uniform rate of growth between 1790 and 1860 and slight but evident changes since the last named date.

One great advantage of the ratio chart is in forecasting. Usually, in business, we forecast by assuming a certain ratio of growth. In the ratio chart we have simply to draw a straight line; usually, in fact, merely to produce the one already drawn representing the rate experienced in the past.

One method by which users of the ordinary chart have attempted to meet its shortcomings in representing statistics of rapid growth is first to draw a "growth axis," or exponential curve [such as that in Fig. 1 *d*, Fig. 2 *d* (curve A), or Fig. 4 *d*] ascending at the *average* rate of growth of the statistics under consideration; then calculate the per cents. of deviation each year from this "growth axis"; and finally, plot these deviations on a separate chart.

This procedure, however, involves, in addition to charting the original statistical figures, much subsidiary calculation and charting; and the results, when obtained, are not as exact as the results obtainable more easily by the ratio method.

Digitized by Google

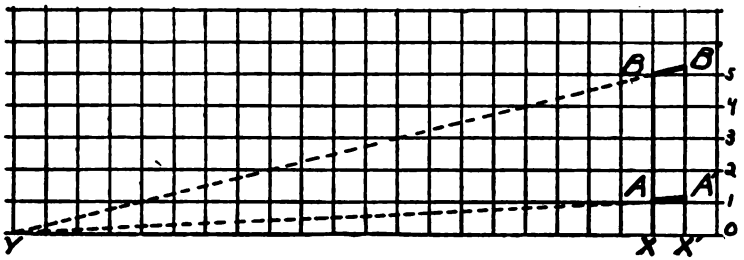


FIG. 6*d*. EQUAL RATES OF GROWTH APPARENTLY UNEQUAL. DIFFERENCE METHOD.
Segments of two curves ascending at equal rates but having unequal slopes and therefore deceiving the eye.

COMPARING CURVES.

If we wish to compare the growth of population in two countries of different sizes, such as Canada and the United States, the ordinary or difference plots, of which we shall assume that AA' and BB' in Fig. 6 *d* are small sections for a given year, will give the impression that the upper curve, *i. e.*, that for the larger country, is ascending at a much faster ratio than the lower, *i. e.*, that for the smaller country. Few would suspect that the two lines AA' and BB' represent precisely the same percentage rate of growth.

In Fig. 6 *r*, on the other hand, the equality of these same two rates of growth is clearly indicated by the parallelism of the two lines AA' and BB'.

Again, if the plots of the two populations appear as in Fig. 7 *d*, most persons, observing the parallelism of the two lines, would jump to the conclusion that the two populations are growing at the same percentage rate. The fact is that the lower line, AA', is ascending at a greater percentage rate than the upper, BB'. The ratio chart (Fig. 7 *r*) shows this at a glance.

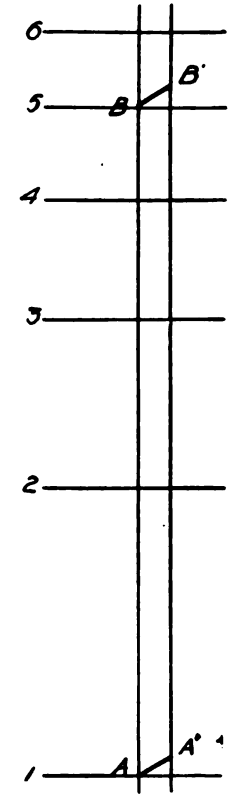


FIG. 6*r*. EQUAL RATES OF GROWTH EVIDENTLY EQUAL. RATIO METHOD.
The two segments are here parallel.

Digitized by Google

In the difference chart, in order to make even a rough eye-estimate of the comparative percentage rates of growth of AA' or BB' (Fig. 6 *d* or Fig. 7 *d*), we must: (1) note the position of

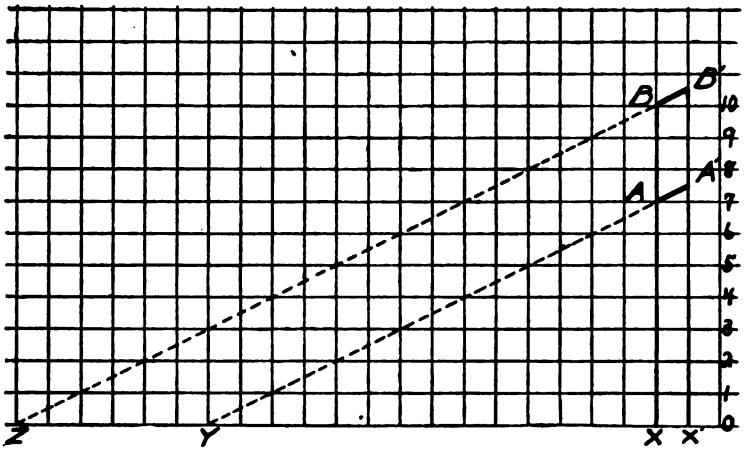


FIG. 7*d*. UNEQUAL RATES OF GROWTH APPARENTLY EQUAL. DIFFERENCE METHOD. Segments of two curves ascending at equal slopes, but at unequal rates and therefore deceiving the eye.

the zero or base-line, XX'; (2) mentally compare $A'X'$ with AX and $B'X'$ with BX; and (3) compare the two comparisons.

Such mental operations are difficult, irksome, and inaccurate, especially if, as is not uncommon, the base or zero line has been omitted in order to economize space. Furthermore they involve shutting our eyes to the slope or steepness of the lines AA' and BB', the very feature which first attracts attention.

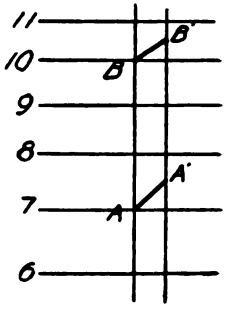


FIG. 7*r*. UNEQUAL RATES OF GROWTH EVIDENTLY UNEQUAL. RATIO METHOD. The two segments are here of different slopes and the steeper slope indicates the greater rate of growth.

Often, in fact, the bottom part of the chart, containing the base line, is cut off, and sometimes, instead of using the same base line for two curves on the same chart the draughtsman will bring one curve nearer the other by using two separate base lines. In either event the result is misleading or confusing and in such cases it is almost hopeless to obtain any clear idea

Digitized by Google

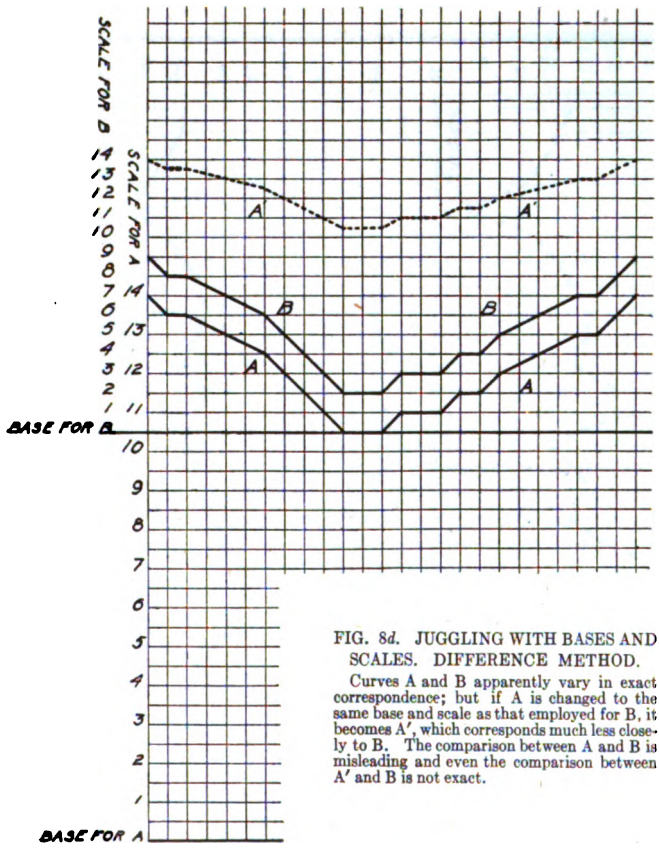


FIG. 8*d*. JUGGLING WITH BASES AND SCALES. DIFFERENCE METHOD.

Curves A and B apparently vary in exact correspondence; but if A is changed to the same base and scale as that employed for B, it becomes A′, which corresponds much less closely to B. The comparison between A and B is misleading and even the comparison between A′ and B is not exact.

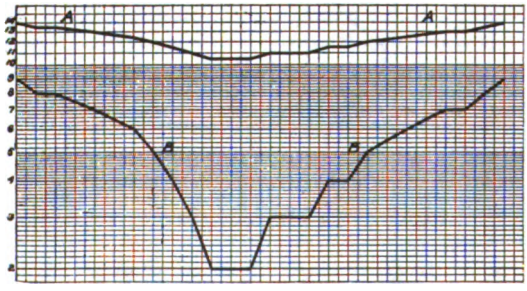


FIG. 8*r*. THE SAME. RATIO METHOD.

The comparison between A and B here exactly represents the facts, showing only a faint resemblance, in marked contrast with the first comparison in the preceding figure.

Digitized by Google

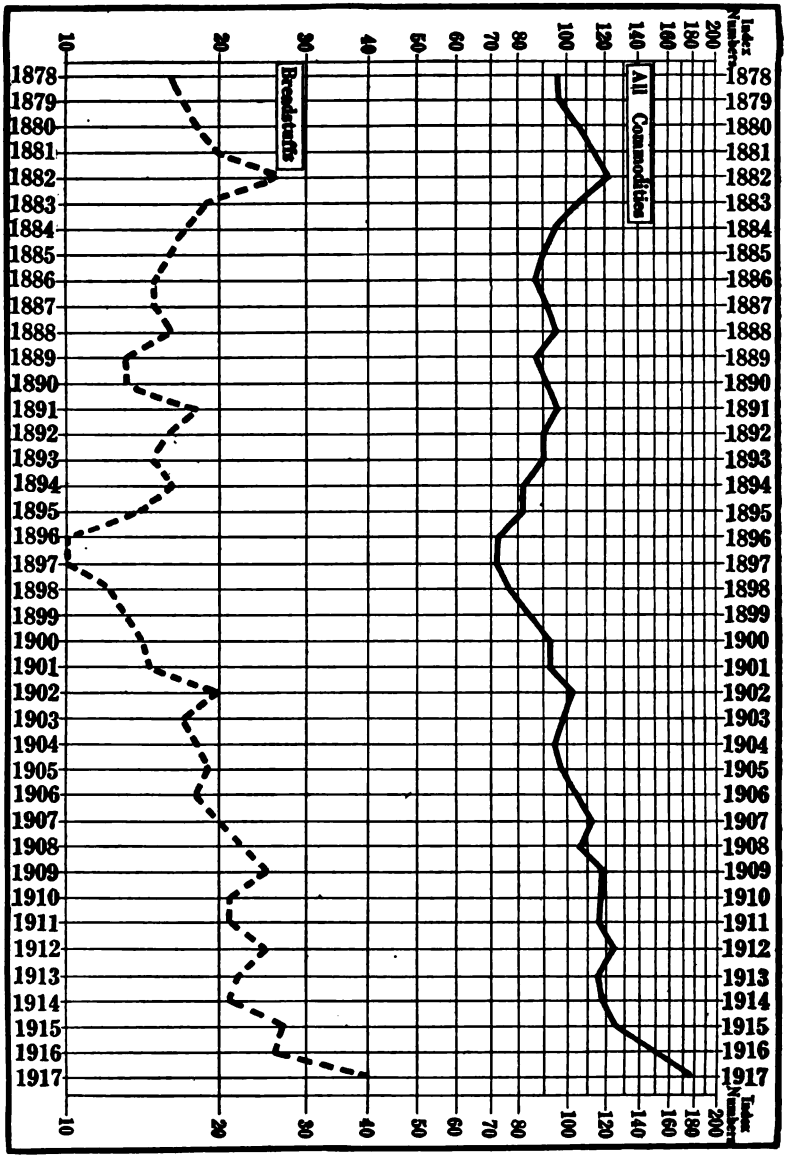


FIG. 9r. THE SAME. RATIO METHOD.

Actually "breadstuffs" fluctuate more than "all commodities."

Digitized by Google

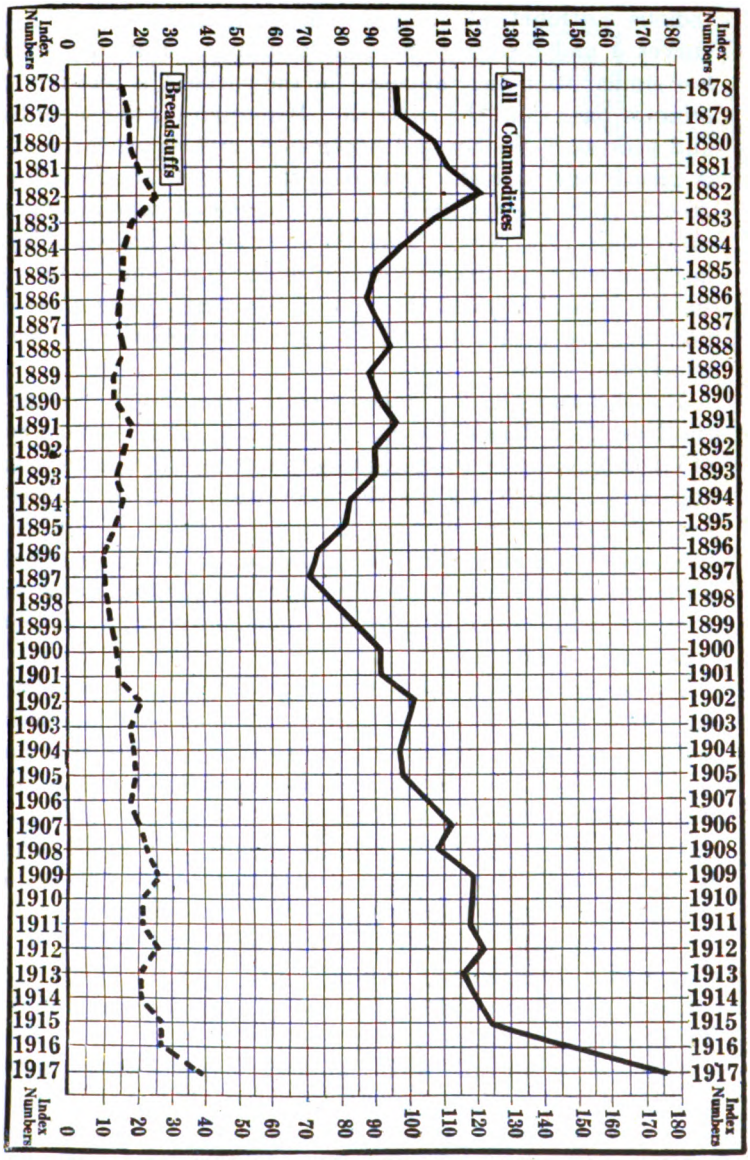


FIG. 94. INDEX NUMBERS COMPARED. DIFFERENCE METHOD.

Apparently "all commodities" fluctuated more than "breadstuffs."

Digitized by Google

of the comparative percentage growths except by recourse to tedious arithmetical computations. Thus in Fig. 8 *d* the curves *A* and *B* seem to be exactly similar. But they are far less similar than they appear; for the curve *A* is relative to a remoter base than the curve *B*. If plotted on the same base and scale as *B*, the curve *A* becomes *A'* and its similarity to *B* is greatly diminished. But even this degree of similarity is greater than the statistics warrant, as we see when using the ratio chart (Fig. 8 *r*). This shows the *exact* degree of similarity of the two curves *A* and *B* which turns out to be very small. Thus the ratio chart is an effective means of avoiding juggling with statistics through base-selection or scale-selection.

Fig. 9 *d* and Fig. 9 *r*, taken from a brief article of mine in the New York *Times Annalist* of March 17, 1917, show a type of error opposite to the foregoing, and one taken from actual statistics. Fig. 9 *d* gives the impression that the prices of breadstuffs have fluctuated less than the prices of "all commodities." On the ratio chart, however, as shown in Fig. 9 *r*, although exactly the same numbers are plotted, it is seen that breadstuffs have actually fluctuated a trifle more violently than "all commodities."

For comparing in detail any two curves on ratio charts, such as those in Fig. 9 *r*, we may do what we have just seen is not properly permissible on a difference chart,—we may move bodily either curve, the upper curve downward or the lower curve upward, until the two are close together. Then the various degrees of parallelism or divergence, at various periods of time, may be seen with the utmost clearness. This is done in Fig. 10 *r*.* Such close comparison will usually give quickly, through the eye, a better practical picture, I think, of the degree of correlation and certainly of the location of the correlation, than can be obtained even by laborious calculations of coefficients of correlation.

Index numbers with widely different bases, such as those of Sauerbeck and of the United States Bureau of Labor, if plotted on *difference* charts without especially selecting the scales or without multiplying or dividing by suitable constants, appear

*The scale on the right applies to the lower curve, that on the left to the upper.

Digitized by Google

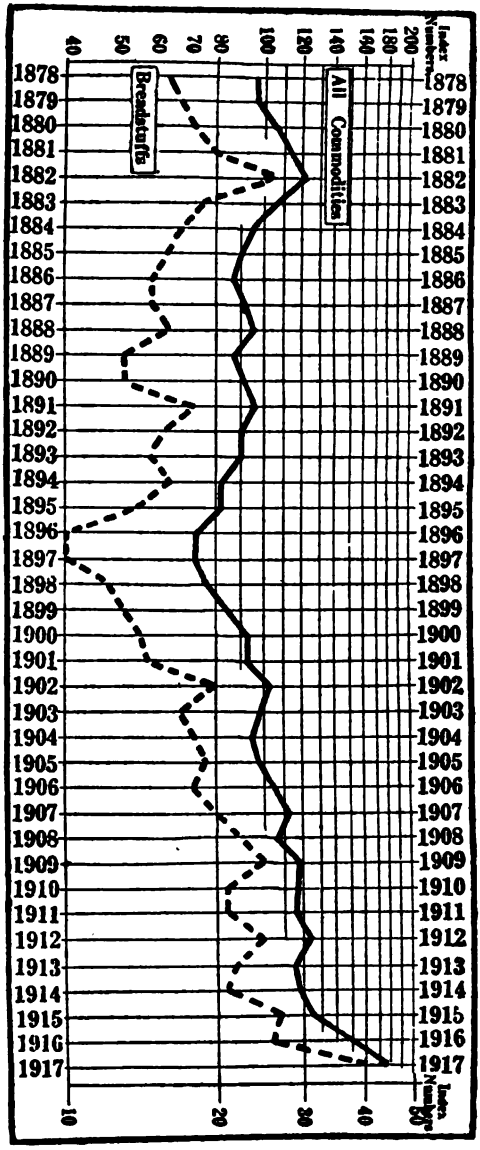


FIG. 10c. THE SAME AS FIG. 9c. RATIO METHOD.

Showing more sharply the contrast by bringing the curves closer together. This procedure is legitimate in the ratio method, but not in the difference method.

Digitized by Google

far more different from each other than they really are. *Ratio* plotting will show their *exact* similarities and differences.

As was said at the outset, difference charts are useful when difference comparisons are wanted. Not only are such cases rare, however, but often difference comparisons between curves are meaningless, and ratio comparisons are alone possible. This is the case when two curves represent incommensurable magnitudes as, for instance, when a curve of railway gross receipts is compared with one of population or one of the circulation of money with one of the price level. In such cases difference plotting involves a very dangerous discretion in selecting the scale and the base in order neither to exaggerate nor to understate the degree of correspondence. For instance, the close similarity in the recent changes in the money in circulation in the United States, on the one hand, and the price level, on the other, was not evident to me until I plotted the two curves on a ratio chart (Fig. 11 *r*). This similarity might be almost overlooked on a difference chart if the scales for dollars and index numbers were selected arbitrarily.

One special, though slight, advantage of the ratio chart is in representing more clearly than in the difference chart the results of multiplying or dividing a statistical series of numbers by a constant, or by another series, and in exhibiting the resultant third series relatively to the other two.*

ATTEMPTS TO MEND THE FAULTS OF THE DIFFERENCE CHART.

One attempt to secure better comparability of curves in ordinary plotting is that customary in handling index numbers, —namely to reduce all prices to percentages of the base-period price, so that we may start all the curves at an even 100 per

*In detail: (1) to multiply or divide by a constant any series of numbers plotted in a curve is merely to shift the curve bodily up or down by the appropriate constant distance; (2) the reciprocals (of a statistical series plotted in a curve) form a second curve symmetrical (relatively to the unity line) to the first curve; in other words two curves representing reciprocal numbers are such as to coincide if the chart is folded over the unity line; thus a plot representing an index number of prices will be the exact reverse of that representing the purchasing power of money, the one ascending whenever the other descends and at the same angle; (3) the curve formed by plotting the products of two series will be such that its distances from the unity line will be the sum of the differences of the other two from that line; (4) analogously, a quotient of two curves is distant from the unity line the difference of their distances (*i. e.*, distance of the dividend curve less that of the divisor curve). Thus if we plot a curve of acres per *capita* of wheat lands and another of bushels per acre and then plot the curve of their products, *i. e.*, of bushels per capita, the three curves will be seen to be related as just described; the same would apply to index numbers of prices, of wages, and of purchasing power of wages ("real wages").

Digitized by Google

cent. If the curves do not greatly diverge this gives approximately correct results. But it is a makeshift method and never gives the absolutely exact comparisons of the ratio

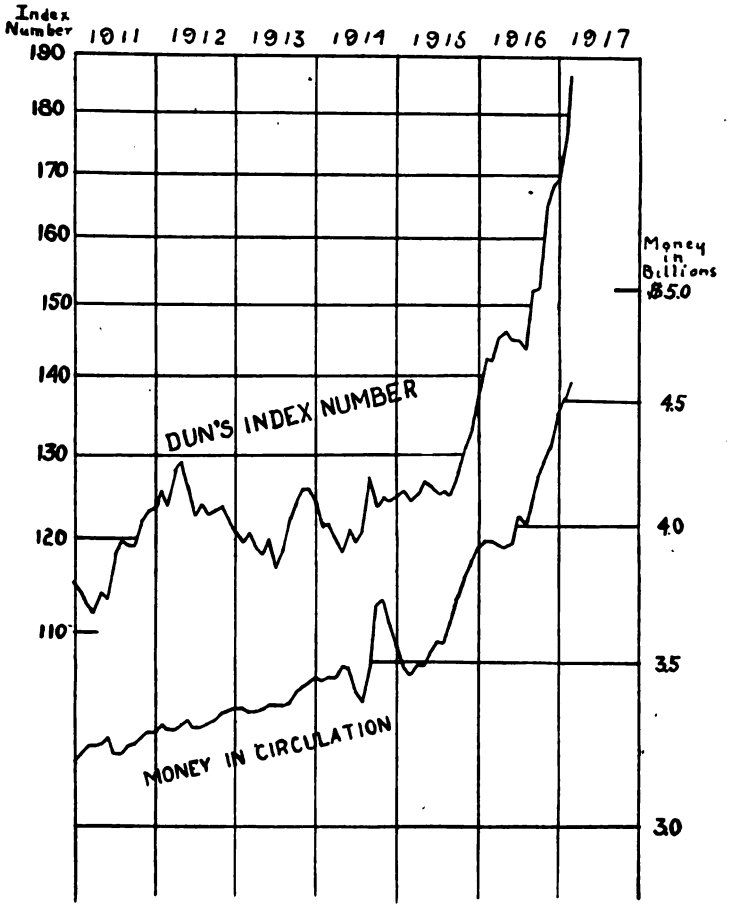


FIG. 11r. MONEY AND THE PRICE LEVEL. RATIO METHOD.

Showing the exact degree in which the price level in the United States has fluctuated in comparison with the amount of money in circulation. Since the war, there has been a close correspondence, changes in the price level following changes in money by two or three months.

method; and, after the curves have diverged considerably, their correct comparison becomes difficult.

Another method is that suggested by Professor Alfred

Digitized by Google

Marshall. It consists in making a special geometric construction.* But recourse to geometric construction is slow and inexact and does not appeal to the eye.

As a matter of fact, users of charts, when seeking percentage comparisons, are apt simply to read off the four numerical figures opposite the various points (as A and A' and B and B' in Fig. 6 *d* or Fig. 7 *d*) and compute arithmetically the two percentage rates of increase, (*i. e.*, the percentage excess of height of A' over A and that of B' over B). But such a procedure is nothing less than giving up the use of the diagram *as* a diagram and using it merely as a table of arithmetical figures. A diagram is supposed to interpret figures to the eye and to need no interpretation itself through arithmetical processes. When it does need such interpretation, it fails of its purpose as a quick, clear, convenient, and reliable picture.

Finally we note that the foregoing correctives and safeguards necessary for using the "difference" method correctly are, in actual fact, almost invariably neglected. Practically no one uses Marshall's subtangent comparison; few even mentally measure the heights of points above the base line,—much less mentally reckon, from these heights, the percentage rates of change; few even notice whether the base line is inserted or omitted or whether there are different base lines for different curves whose inconsistency needs to be allowed for; growth axes are seldom constructed; few make any effort to shut their eyes to straightness or parallelism in order to avoid being misled into assuming uniformity or similarity of *percentage* changes; few, even, use a diagram as a table of figures.

Professor Marshall says of the ordinary method of curve-plotting: "Its defects are such that many statisticians seldom use it except for the purpose of popular exposition, and for this purpose, I must confess, it has great dangers."† A business

* "On the Graphic Method of Statistics," Jubilee Vol. of the Royal Statistical Society, 1885, pp. 251–60. To help translate the deceptive slope of different lines, such as AA' and BB' in Fig. 6 *d* or Fig. 7 *d*, we produce these lines so that they cut the base line. If, when so produced, they are found to cut the base line at the same point, as is the case in Fig. 6 *d*, where they both cut at Y, then we may know that they have the same percentage slope. If, as in Fig. 7 *d*, they do not cut the base line in the same point, the one which cuts the further off has the smaller percentage rate of increase and the rates of increase per annum of the two lines at A and B are inversely proportional to the distances at which they cut the base as measured from X. Thus if, in Fig. 7 *d*, XY is 14/20 of XZ the percentage change at A is 20/14 of that at B. Again, in Fig. 4 *d* the slopes at A and A' are equal, as $XY = X'Y'$. All of these results may be proved geometrically.

† "On the Graphic Method of Statistics," Journal of the Royal Statistical Society, 1885, p. 251.

Digitized by Google

man whose attention was recently called to the advantages of the ratio chart recharted his business statistics and was startled to discover how he had been misled by the ordinary method.

After such a severe, but, I believe, just, indictment against ordinary or "difference" plotting it may be asked whether, as usually employed, such plotting is not worse than useless. It is certainly true that ordinarily it misleads to some extent. I have sometimes found myself hesitating to use a curve in a publication, because, on the one hand, of the space it would take up if the base line were inserted and, on the other, of the mistaken impression it might make if the base line were omitted.

The best that can be said for the difference method is: it always shows whether there is an increase or decrease; it usually displays the grosser contrasts at a glance; the base or zero line gives a means, lacking in the ratio method, for plotting zeros, for comparing positive and negative quantities, and for seeing in a simple and self-evident comparison the vertical elevations of points in a curve above the base line.

WAYS OF UTILIZING DIRECTION IN THE RATIO CHART.

The eye reads a ratio chart more rapidly than a difference chart or a table of figures. We may recapitulate what most easily catches the eye as follows:

1. If we see a curve ascending, and nearly straight, we know that the statistical magnitude it represents is increasing at a nearly uniform rate.
2. If the curve is descending, and nearly straight, the statistical magnitude is decreasing at a nearly uniform rate.
3. If the curve bends upward the rate of growth is increasing.
4. If downward, decreasing.
5. If the direction of the curve in one portion is the same as in some other portion it indicates the same percentage rate of change in both.
6. If the curve is steeper in one portion than in another portion it indicates a more rapid rate of change in the former than in the latter.
7. If two curves on the same ratio chart run parallel they represent equal percentage rates of change.
8. If one is steeper than another the first is changing at a faster percentage rate than the second.

Digitized by Google

9. The imaginary straight line most nearly representing, to the eye, the general trend of the curve, is its "growth axis," and represents the average rate of increase (or decrease); and the deviations of the curve from this growth axis are plainly evident without recharting.

10. The slope of the imaginary line between any two points on a curve indicates the average rate of change between the two.

WAYS OF UTILIZING ELEVATION IN THE RATIO CHART.

The preceding relates to *direction*. As to *elevation*, the eye can, with a little familiarity, translate vertical elevation into numerical ratio; for, as we have seen, a certain elevation represents a 10 per cent. increase, another a 100 per cent. increase, or doubling, etc., etc.*

To one accustomed to using the difference chart the use of elevations on the ratio chart may at first be confusing. But only a few minutes are necessary to learn its use; and this use of the mere elevation for measuring the ratios, or vertical distances, between any two points between the two magnitudes represented by those points, is really easier and more exact than the more self-evident method, in the ordinary chart, of comparing their two vertical distances from the base line.†

* For roughly interpreting elevations without having to read the marginal figures or measure distances, the original system of equidistant horisontals, as shown in Fig. 1 *r*, is not inconvenient. The chart is divided into horisontal 10 per cent. bands and the eye can approximately reckon how far any point on a statistical curve is vertically above any other point,—whether one band's breadth, or two, or any fractional part, and, therefore, whether it represents a magnitude 10 per cent. more, or 10 per cent. *more than* 10 per cent. more, *i. e.* 21 per cent. more, etc.

†We may also, if we wish, make much more exact comparisons than those afforded by the glance of an eye. We may exactly measure any slope such as that of a tangent to a curve at any point or such as that of a line connecting any two points on the curve. This may be done in accordance with the constructions of the radiating lines in Fig. 3 *r;* by drawing a line parellel to that of which the slope is desired, from any convenient point in the horisontal line "10". Let this cut the vertical line one unit to the right of said point; the height of this intersection above (or below) line "10" measures exactly the rate of ascent (or descent) of the slope to be measured.

Again we may exactly measure any ratio comparison between two points on the same curve—or on different curves for that matter. All we need to know is the vertical distance between the two points and compare this with the vertical scale above (or below) the line "10," as illustrated in Fig. 3 *r*. Such a comparison may be made by a draughtsman's dividers, parallel rulers, sliding triangle, or even by using the edge of a sheet of paper and a lead pencil. Thus, if the elevation, as measured, is found equal to the distance between "1" and "2" (or "10" and "20" or "100" and "200") then the upper point represents a statistical magnitude just twice that represented by the lower point; or again, if the measured elevation is equal to the distance between "1" and "1.37" (or "10" and "13.7" or "100" and "137") on the scale, then the upper point represents a statistical magnitude 37 per cent. greater than that represented by the lower point.

Digitized by Google

SUMMARY.

In the ratio method, then, a straight line* always represents a constant percentage rate of increase or decrease and, conversely, a constant percentage rate of increase or decrease is always represented by a straight line; a curve deviating from a straight line invariably implies that the percentage of change deviates correspondingly from constancy; any two curves or two portions of the same curve which are parallel represent exactly equal percentage rates of change; any two curves or portions of curves which show a contrast of direction always indicate a corresponding contrast in percentage change; if the numbers plotted are halved or changed in any other ratio, the resulting curve will simply be raised or lowered but will maintain exactly the same series of directions and therefore present the same appearance to the eye; if the scale is properly selected, a curve is never nearly horizontal except when it actually represents an almost infinitesimal rate of increase or decrease, nor is it ever nearly vertical except when it actually represents a rate correspondingly enormous; as there is no zero line there is no waste space on its account and the diagrams can be cut off close, both above and below the curve; there can be no juggling with base lines or scales; there is no need of special supplementary geometric constructions, such as Marshall's subtangent construction; there is no need of laborious calculations to reduce original figures to index numbers or percentages; there is no need of eliminating the growth axis (which, in the ratio method, is simply a straight line, the percentage deviations from which are apparent without special calculation or replotting).

The features of a curve which, whether we will or not, most "catch" the eye are concerned with comparative *direction*,—straightness or curvedness; steepness or flatness; parallelism or divergence. These features therefore ought to be, not

*It is interesting to note that engineers have found it advantageous to devise special plotting charts which will reduce parabolic, hyperbolic and probability curves of the ordinary charts to straight lines on the special charts, *e. g.*, see A. S. Langsdorf, "Methods for Determining the Equations of Experimental Curves," Journal of the Association of Engineering Societies, June, 1904, pp. 325-43; L. F. Harza, "Notes on Determination of Experimental Equations," Wisconsin Engineer, December, 1906; George C. Whipple, "Element of Chance in Sanitation," Journal of the Franklin Institute, Vol. 182, July 7, 1916, p. 37; Allen Hasen, "Storage to be Provided in Impounding Reservoirs for Municipal Water Supply," Proceedings of the American Society of Civil Engineers, Nov., 1913.

Digitized by Google

snares nor stumbling blocks, as they are in the "difference" chart, but aids or sign boards as they are in the "ratio" chart.

And, besides the full utilization of *direction*, we have, in the ratio chart, that of *elevation*. While the interpretation by the eye and mind of elevation requires a little preliminary training it soon becomes easier, more rapid and more accurate than the corresponding procedure for difference charts.

In a word, the ratio chart simply utilizes the natural powers of the eye. Consequently, when one is once accustomed to it, it never misleads, but always pictures a multitude of ratio relations at a glance, with absolute fidelity and without the annoyance of reservations or corrections.

BIBLIOGRAPHY.

Professor Jevons seems to have been the first* to employ the idea of the ratio chart. He did not use prepared ruling but employed a table of logarithms and plotted the logarithms of the statistical numbers. While this procedure amounts to the same thing it fails to interpret the method to the average man, for it seems to lose touch with the actual statistical figures.

W. S. JEVONS, *Investigations in Currency and Finance*, London, (Macmillan) 1884, pp. 53, 128, 150.

ALFRED MARSHALL, *On the Graphic Method of Statistics*, Jubilee Volume of the Royal Statistical Society, June 22-24, 1885, p. 258.

GEORGE H. WOOD, *Some Statistics relating to Working Class Progress since 1860*, Journal of the Royal Statistical Society, December, 1899, p. 660.

ARTHUR L. BOWLEY, *Elements of Statistics*, London (King), 1901, pp. 188–196.

R. BENINI, *Principii di Statistica Metodologica*, Torino (Unione Tipografico-Editriee Torinese), 1906, pp. 144–148.

W. J. CUNNINGHAM, *The Logarithmic Scale in Graphic Charts*, Railroad Age Gazette, June 25, 1909, pp. 1517–19.

WALTER T. RAY, *Rational Stock Speculation*, Norfolk, Va. (published by the author), 1914, Vol. 1, pp. 15–20.

*"During the years" (viz., 1863 to 1884) "which have elapsed since this paper and previous essays on the value of gold were printed, I have not happened to learn that logarithmic diagrams had been used before their employment in these inquiries. It must be evident, however, that they furnish the true mode of representing all statistical and other numbers of which the ratios, not the absolute amounts, are in question. Now this is, strictly speaking, the case with almost all numbers used for statistical research, as apart from mere practical accounts." W. S. Jevons, "Investigations in Currency and Finance," London, (Macmillan), 1884, p. 128.

Digitized by Google

WILLARD C. BRINTON, *Graphic Methods for Presenting Facts*, New York (Engineering Magazine Company), 1914, pp. 132–7.

WILLARD C. BRINTON and others, *Preliminary Report of the Joint Committee on Standards for Graphic Presentation*, 1915, p. 5. Printed by the American Society of Mechanical Engineers.

IRVING FISHER, *Plotting Ratios Along with Amounts*, New York Times Annalist, March 19, 1917, p. 398.

JOHN WENZEL, *Graphic Charts*, Scientific American Supplement, No. 2154, April 14, 1917, p. 236.

Digitized by Google

ON THE VARIATE DIFFERENCE CORRELATION METHOD AND CURVE-FITTING.

By Warren M. Persons, *Harvard University.*

(Professors A. A. Young of Cornell University and E. E. Day of Harvard University have generously read the manuscript of this paper and have made helpful suggestions. Computations and pertinent comments have been made by Mr. Edwin Frickey of Colorado College. W. M. P.)

I.

In *Biometrika* for April, 1914, "Student" generalized the method of differences for the elimination of spurious correlation due to order of items in time or space. In the same *Journal* for November, 1914, Dr. O. Anderson of Petrograd provided the probable errors of the successive difference correlations of two series where the correlations of random pairs of the variates are zero. In the same number B. M. Cave and Karl Pearson presented "Numerical Illustrations of the Variate Difference Correlation Method," using Italian economic data.

The conclusion of "Student" is this: "If we wish to eliminate variability due to position in time or space and to determine whether there is any correlation between the residual variations, all that has to be done is to correlate the 1st, 2nd, 3rd. . . nth differences between successive values of our variable with the 1st, 2nd, 3rd . . . nth differences between successive values of the other variable. When the correlation between the two nth differences is equal to that between the two $(n+1)$th differences, this value gives the correlation required."* The meaning of "Student" is that the correlation required is indicated by the ultimate steadiness of values of the correlation coefficient for higher multiple differences of the items.†

After working with 11 series of 28 items each Miss Cave and Professor Pearson stated:‡ "In most cases our difference cor-

* *Biometrika*, April, 1914, p. 179.
† *Biometrika*, November, 1914, footnote p. 340.
‡ *Ibid.*, pp. 354–355.

Digitized by Google

relations have hardly even with the sixth differences reached a steady state. . . . In the great bulk of instances there is still a more or less steady rising or falling appreciable in the difference correlations, and all we can really say is that the final value, the true r_{XY}, will be somewhat greater or less than a given number. From an examination of the actual numerical working of the correlations, it appears to us that the terminal values are in the case of these short series of very great importance. It is further clear that the theory as given by 'Student' depends upon certain equalities which are not fulfilled in practice in short series. We await with much interest the complete publication of Dr. Anderson's work, and hope to find a fuller discussion of the allowance to be made in short series for the influence of the terminal state of affairs on the steadiness of the series and on the approach to the standard deviation formulae. But apart from these lesser points, our present numerical investigation has convinced us of the very great value of the new method of Variate Difference Correlations."

In the demonstration leading to his theorem, previously quoted, "Student" stated that "if x_1, x_2, x_3, etc., y_1, y_2, y_3, etc., be corresponding values of the variables x and y, then if x_1, x_2, x_3, etc., y_1, y_2, y_3, etc., are randomly distributed in time and space, it is easy to show that the correlation between the corresponding nth differences is the same as that between x and y."* In the proof of this statement certain assumptions were made as follows: first, that there is a significant correlation between items with similar subscripts, *i. e.*, $\Sigma x_K y_K \neq 0$; second, that if the items of each series be paired with the preceding items of the same series the correlation will be zero, *i. e.*, $\Sigma(x_K x_{K+1}) = 0$ and $\Sigma y_K y_{K+1} = 0$; third, that there is no significant correlation between the items of one series and the items of the other when there is a lag in either direction, *i. e.*, $\Sigma(x_K y_{K+1}) = 0$, $\Sigma(x_{K+1} y_K) = 0$; fourth, that the sum of the items of each series is zero, *i. e.*, $\Sigma x_K = 0$, $\Sigma y_K = 0$; fifth, that the time element in any series ordered in time can be expressed as an algebraic function of time (t) of some degree, the nth.

* *Biometrika*, April, 1914, p. 179.

Digitized by Google

It is my contention that these assumptions are such as cannot be retained in applying the method to the most common types of problems. For instance the pairing of items of two time series is made possible by the position of those items in time either because they occur in the same time interval (concurrent) or in definitely related intervals (lag). Our problem may be, and usually is, not only to determine the correlation but to find what pairings give the maximum correlation. In such case the assumption that only one pairing is significant vitiates the conclusion at the outset. The writers on the variate difference correlation method all assume that "the true r_{XY}" is for pairs concurrent in time.

Let the method of variate differences be applied to two time series artificially constructed. Let series *A* be made up of successive values of the function $2t+3$ with $+1$ and -2 alternately added to the items. Let series *B* be made up of the values of successive values of the function t^2+3t+2 with -1 and $+1$ alternately added to the items. We will have the following series and differences:

SERIES A.

t	*Items*	*1st D*	*2nd D*	*3rd D*
0	4	..	..	..
1	3	−1	..	..
2	8	+5	+6	..
3	7	−1	−6	−12
4	12	+5	+6	+12
5	11	−1	−6	−12
		etc.		

SERIES B.

t	*Items*	*1st D*	*2nd D*	*3rd D*
0	1	..	..	..
1	7	+6	..	..
2	11	+4	−2	..
3	21	+10	+6	+8
4	29	+8	−2	−8
5	43	+14	+6	+8
		etc.		

Digitized by Google

The process of taking differences eliminates the elements due respectively to the first and second degree functions of time. The oscillating elements remain. If concurrent items for the third and higher differences are paired, we have $r_o^{iii} = r_o^{iv} = \ldots r_o^{(m)} = -1$. If the items are paired with a lag of one in either direction we have $r_{-1}^{iii} = r_{+1}^{iii} = r_{-1}^{iv} = r_{+1}^{iv} = \ldots r_{-1}^{(m)} = r_{+1}^{(m)} = +1$.* In this case the time elements are eliminated as far as they can be by the method of differences and yet the series resulting are not, properly speaking, random in time. It will be of interest to determine if this sort of oscillation occurs in actual data.

II.

I have recently been working with some 21 series of economic statistics for the United States for the period 1879 to 1913, giving 35 items to each series. Application of the variate difference correlation method to such series has forced me to the conclusion that neither the possibilities nor the limitations of the method, when applied to short series, have been appreciated by the writers on the subject.†

The applications of the method by "Student" and by Miss Cave and Professor Pearson are not satisfactory tests: first, because they have applied it merely to items of the same date, thus assuming that the real correlation can exist only for such pairs; and, second, they assume that the correlation indicated by the steadiness of coefficients between higher differences only is significant, the coefficient for first differences not being significant unless it is supported by steadiness of the coefficients of higher differences. My objections to these assumptions and the conclusions based upon them will be il-

* In r_t the subscript t indicates the number of time intervals that the items of series A precede (−) or lag behind (+) the corresponding items of series B.

† At the conclusion of his article in *Biometrika*, April, 1914, pp. 269–279, Mr. O. Anderson makes the following statement, which I translate:

"If we take into consideration that for our purposes the evolutionary component of a series has disappeared if it becomes so small relative to the oscillatory component that it can influence only the 3rd, 4th, etc., decimal place of the expression for R [the coefficient of correlation] then we may conclude that not only components which are represented by a parabola of higher order, but also those represented by transcendental functions (such as a sine curve) become eliminated taking by a finite number of differences. Further it may be shown that generally all more or less 'smooth series' all of which are characterised by a considerable degree of positive correlation between adjacent items, lose the character of smoothness in the process of multiple differences. The generalised Cave-Hooker procedure is, therefore, manifestly a quite universal means of sifting out the correlation between the oscillatory elements of complex series."

Digitized by Google

lustrated by application of the variate difference correlation method to American data for the period 1879–1913. The series used are the following:

1. Wholesale prices of commodities.
2. Gross receipts of railroads.
3. Net earnings of railroads.
4. Coal production.
5. Exports from the United States.
6. Imports into the United States.
7. Pig-iron production.
8. Price of pig-iron.
9. Immigration (fiscal year).
10. Shares sold on the New York Stock Exchange.
11. Average price of shares sold on the New York Stock Exchange.
12. New York clearings plus five times outside clearings (called clearing index).
13. Clearing index divided by relative wholesale prices (called corrected clearing index).
14. New railroad mileage constructed.
15. Per cent. of business failures.
16. Liabilities of business failures.
17. Balance of trade.
18. Weighted index numbers of the yield per acre of nine leading crops.
19. Ratio of loans to resources of banks.
20. Ratio of cash to deposits of banks.
21. Surplus reserves of New York associated banks.

In each case, except for the clearing index, I assumed the secular trend to be linear. A straight line was fitted to each series by the method of least squares or the method of moments. For the clearing index I assumed the secular trend to be the compound interest law and the function $y = B\,C^t$, where t represents time and B and C are constants determined by the data, was fitted to the series. The deviations of the raw figures from the lines of secular trend were found and designated the "cycles."

Digitized by Google

TABLE I.

COEFFICIENTS OF CORRELATION FOR THE CYCLES OF THE BUSINESS BAROMETER AND THE CYCLES OF WHOLESALE PRICES, GROSS RECEIPTS OF RAILROADS, CORRECTED CLEARING INDEX AND SURPLUS RESERVES OF NEW YORK BANKS TOGETHER WITH COEFFICIENTS FOR MULTIPLE DIFFERENCES, FIRST TO SIXTH, WITH VARIOUS DEGREES OF LAG, 1879–1913.

Business Barometer Correlated with.	Coefficients of Correlation. (a)						
	r_{-3}.	r_{-2}.	r_{-1}.	r_0.	r_{+1}.	r_{+2}.	r_{+3}.
Wholesale prices, Cycles	....	....	+.90	+.95	+.85	....	....
1st D.	....	....	+.20	+.78	−.03	....	....
2nd D.	....	....	−.22	+.78	−.46	....	....
3rd D.	....	....	−.35	+.78	−.57	....	....
4th D.	....	....	−.40	+.76	−.59	....	....
5th D.	+.12	−.01	−.42	+.74	−.58	+.15	+.09
6th D.	....	....	....	+.71	....	....	....
Gross Receipts of Railroads							
Cycles	....	....	+.91	+.94	+.79	....	....
1st D.	....	....	+.34	+.80	−.28	....	....
2nd D.	....	....	+.08	+.74	−.64	....	....
3rd D.	....	....	−.06	+.74	−.71	....	....
4th D.	....	....	−.18	+.76	−.74	....	....
5th D.	+.25	−.22	−.24	+.75	−.75	+.25	+.09
6th D.	....	....	....	+.74	....	....	....
Corrected Clearing Index							
Cycles	....	....	+.51	+.77	+.80	+.66	....
1st D.	....	....	−.35	+.32	+.52	−.04	....
2nd D.	....	....	−.55	+.23	+.38	−.29	....
3rd D.	....	....	−.55	+.21	+.27	−.25	....
4th D.	....	....	−.53	+.21	+.18	−.18	....
5th D.	−.36	+.49	−.53	+.24	+.07	−.07	−.20
6th D.	−.48	+.53	−.52	+.28	−.05	+.05	....
Surplus Reserves of New York Banks							
Cycles	....	....	−.27	−.60	−.62	....	....
1st D.	....	....	+.33	−.72	+.02	....	....
2nd D.	....	....	+.56	−.74	+.23	....	....
3rd D.	....	....	+.64	−.75	+.35	....	....
4th D.	....	....	+.71	−.77	+.42	....	....
5th D.	+.28	−.46	+.75	−.78	+.49	−.20	+.17
6th D.	....	....	+.79	−.81	+.59	....	....

(a) The subscript i of the coefficient of correlation r indicates the years lag (+) or years previous (−) of the Business Barometer.

Investigation of the 21 series of cycles led to the conclusion that the fluctuations of 9 of them synchronize and hence can logically be combined into a business barometer. Consequently a business barometer is constructed of the 9 series, those numbered 1 to 9 in the list given on the preceding page.* The coefficients of correlation for the cycles of the business barometer and the cycles of wholesale prices, gross receipts of railroads, corrected clearing index and surplus reserves of

* For details of this investigation see the article by the writer in the *American Economic Review*, December, 1916.

Digitized by Google

New York associated banks, together with the coefficients of correlation for the first to sixth differences, with various degrees of lag for cycles and differences, are given in Table I. Of course there is an element of spurious correlation in the coefficients for the business barometer, wholesale prices, and gross receipts of railroads because the two last named series enter into the barometer. That element is not believed to be large for the cycles and first differences.

Examination of Table I reveals: first, high, positive, and steady coefficients for concurrent items for wholesale prices and gross receipts of railroads, high, negative, and steady coefficients for surplus reserves, low, positive, and fairly steady coefficients for the corrected clearing index; second, all the coefficients for higher differences show a marked tendency to alternate in algebraic sign as successive degrees of lag are taken in either direction.

TABLE II.

PROBABLE ERRORS OF COEFFICIENTS OF CORRELATION FOR SERIES OF 35 AND 46 ITEMS, RESPECTIVELY, AND FOR THEIR MULTIPLE DIFFERENCES, FIRST TO SIXTH.(a)

(Thirty-five Items in Original Series.)

Coefficients of Correlation.	Items.	1st D.	2nd D.	3rd D.	4th D.	5th D.	6th D.
.90	.02	.03	.03	.03	.04	.04	.04
.80	.04	.05	.06	.07	.07	.08	.08
.70	.06	.07	.08	.09	.10	.11	.11
.60	.07	.09	.10	.12	.12	.13	.14
.50	.09	.11	.12	.13	.15	.16	.16
.40	.10	.12	.14	.15	.16	.17	.18
.30	.10	.13	.15	.16	.18	.19	.20
.20	.11	.14	.16	.17	.19	.20	.21
.10	.11	.14	.16	.18	.19	.21	.22
.00	.11	.14	.16	.18	.19	.21	.22

(Forty-six Items in Original Series.)

.90	.02	.02	.03	.03	.03	.03	.04
.80	.04	.04	.05	.06	.06	.06	.07
.70	.05	.06	.07	.08	.09	.09	.10
.60	.06	.08	.09	.10	.11	.11	.12
.50	.08	.09	.11	.12	.13	.13	.14
.40	.08	.10	.12	.13	.14	.15	.16
.30	.09	.11	.13	.14	.15	.16	.17
.20	.10	.12	.14	.15	.16	.17	.18
.10	.10	.12	.14	.15	.17	.18	.19
.00	.10	.12	.14	.16	.17	.18	.19

(a) Computed from formulae developed by O. Anderson and A. Ritchie-Scott. See *Biometrika*, November, 1915, p. 136.

Digitized by Google

The significance of the various coefficients depends not only on their size but upon the number of items used in the computation. The probable errors for various coefficients based on 35 items in the original series, first to sixth differences are given in Table II.* The probable errors for sixth differences are, approximately, twice those for the original series. A coefficient of, say .45 or more would be significant for the original series and of .65 or more for sixth differences. What is the explanation of the observed steadiness, and of the alternation of sign of coefficients for various degrees of lag? "Student" believes that the steadiness is due to the random distribution, with respect to time, of the differences. The alternation in sign is a phenomenon not noticed, or if noticed not considered, by the writers on the subject.

First, let us consider the phenomenon of alternation in sign. Let one of the original series be $x_0, x_1, x_2 \ . \ . \ . \ x_{n-1}$, the first differences being $x_1-x_0, x_2-x_1, \ . \ . \ . \ x_{n-1}-x_{n-2}$.† Suppose the first differences alternate in sign at any point so that we have

$$x_K-x_{K-1}=+a$$
$$x_{K+1}-x_K=-b$$
$$x_{K+2}-x_{K+1}=+c$$
$$\text{etc.}$$

where a, b, c . . . are positive numbers. The second differences are $-b-a, +c+b, -d-c, \ . \ . \ .$, a series alternating in sign and *larger* numerically than the series from which it is derived. For the portion of the original series, however, where there is no alternation in sign the first differences will be *smaller* numerically than the items from which they are derived. Since the nth differences are derived from the $(n-1)$th differences in the same way that the first differences are derived from the original series we have the following conclusion: If consecutive items of a series alternate in sign the first and higher differences will also alternate in sign and the resulting items will increase numerically as the order of

* These probable errors are computed from theories developed by O. Anderson, *Biometrika*, April, 1914, p. 269. The formulae for probable errors *may* hold for a very large number of items but I doubt their validity for less than 100 items.

† The standard notation for successive finite differences is $\Delta, \Delta^2, \Delta^3$, etc. That notation is not used in this paper because it would tend to conceal relations brought out in Part III of this paper.

Digitized by Google

the difference increases. A succession of like signs may persist with the first and higher differences but the numbers resulting will be smaller numerically than those resulting where the sign alternates. Where the variate difference method is applied to two short series we may, therefore, expect the terms alternating in size to be of dominating influence upon the coefficient of correlation. Also when a lag is taken in either direction the coefficients will tend to alternate in sign.

TABLE III.

PERCENTAGES FOUND BY TAKING THE RATIO OF THE NUMBER OF CASES IN WHICH SUCCESSIVE ITEMS DIFFER IN SIGN TO THE POSSIBLE NUMBER OF ALTERNATIONS IN SIGN OF SUCCESSIVE ITEMS; VARIOUS SERIES, FIRST TO SIXTH DIFFERENCES, 1879–1913.

Series.	Percentage of Unlike Signs. (a)					
	1st D.	2nd D.	3rd D.	4th D.	5th D.	6th D.
Business barometer	.55	.59	.64	.87	.79	.79
Wholesale prices	.42	.59	.61	.63	.83	.86
Liabilities of commercial failures	.58	.62	.68	.70	.69	.75
Gross receipts of railroads	.33	.66	.68	.70	.76	.82
Percentage of loans to resources of banks	.55	.66	.68	.73	.72	.82
Shares sold on New York Stock Exchange	.58	.62	.71	.70	.79	.82
Immigration	.42	.59	.61	.67	.72	.79
Coal produced	.52	.69	.74	.83	.93	.93
Percentage of cash to deposits in banks	.52	.78	.84	.83	.86	.89
Average price of shares on N. Y. St. Exch.	.55	.75	.74	.77	.76	.86
Corrected Clearing index	.61	.62	.71	.77	.79	.79
Exports of merchandise	.42	.62	.68	.80	.83	.82
Imports of merchandise	.58	.66	.68	.80	.79	.82
Pig-iron produced	.55	.62	.61	.60	.59	.71
Net earnings of railroads	.30	.53	.68	.70	.83	.82
Price of pig-iron	.55	.62	.61	.73	.79	.79
Surplus reserves of New York banks	.45	.66	.64	.70	.69	.71
New railroad mileage constructed	.55	.66	.81	.83	.83	.82
Indices of crop yield, 3-yr. averages	.48	.66	.77	.93	.93	.93
Average of percentages	.50	.64	.70	.75	.79	.82

(a) A + or − item followed by a null item or *vice versa* is counted ½. The possible numbers of unlike signs are 33, 32, 31, 30, 29 and 28 for the differences, 1st to 6th.

Table III shows the tendency of the higher differences of the series here under consideration to alternate in sign. With 34 items in the series of first differences there are 33 possible alternations in sign when each term is compared with the preceding term; there are also 33 possible cases of steadiness in sign. Counting the number of times that the signs of successive terms of each series alternate and expressing the number as a percentage of the possible number (32 for second differences, 31 for third differences, etc.) we have the percentages appearing in the table. For numbers chosen at

Digitized by Google

TABLE IV.

COEFFICIENTS OF CORRELATION (a) BETWEEN CONCURRENT ITEMS AND THOSE HAVING A LAG OF ONE IN EITHER DIRECTION, FIRST TO FOURTH DIFFERENCE, OF ALL POSSIBLE COMBINATIONS OF THE FOLLOWING SERIES FOR THE PERIOD 1879–1913:

A. Business Barometer.
B. Wholesale Prices.
C. Gross Receipts of Railroads.
D. Corrected Clearing Index.
E. Surplus Reserves of New York Banks.

	AA.			BB.			CC.			DD.			EE.		
Difference.	$r_{-1.}$	$r_{0.}$	$r_{+1.}$	$r_{-1.}$	$r_{0.}$	$r_{+1.}$	$r_{-1.}$	$r_{0.}$	$r_{+1.}$	$r_{-1.}$	$r_{0.}$	$r_{+1.}$	$r_{-1.}$	$r_{0.}$	$r_{+1.}$
First	−.13	+1.00	−.13	+.24	+1.00	+.24	+.05	+1.00	+.05	−.31	+1.00	−.31	−.33	+1.00	−.33
Second	−.49	+1.00	−.49	−.43	+1.00	−.43	−.33	+1.00	−.33	−.54	+1.00	−.54	−.54	+1.00	−.54
Third	−.61	+1.00	−.61	−.64	+1.00	−.64	−.44	+1.00	−.44	−.66	+1.00	−.66	−.65	+1.00	−.65
Fourth	−.69	+1.00	−.69	−.71	+1.00	−.71	−.54	+1.00	−.54	−.74	+1.00	−.74	−.71	+1.00	−.71
	AB.			AC.			AD.			AE.			BC.		
First	+.20	+.78	−.03	+.34	+.80	−.28	−.35	+.32	+.52	+.33	−.72	+.02	+.27	+.66	+.07
Second	−.22	+.78	−.46	+.08	+.74	−.64	−.55	+.23	+.38	+.56	−.74	+.23	−.06	+.62	−.40
Third	−.35	+.78	−.57	−.06	+.74	−.71	−.55	+.21	+.27	+.64	−.75	+.35	−.18	+.64	−.52
Fourth	−.40	+.76	−.59	−.18	+.76	−.74	−.53	+.21	+.18	+.71	−.77	+.42	−.26	+.68	−.58
	BD.			BE.			CD.			CE.			DE.		
First	−.10	+.04	+.66	+.19	−.50	−.21	−.34	+.06	+.59	+.41	−.49	−.27	−.27	−.47	+.43
Second	−.17	−.23	+.67	+.45	−.52	+.03	−.42	−.06	+.50	+.59	−.51	−.12	−.11	−.42	+.51
Third	−.14	−.30	+.65	+.51	−.53	+.15	−.36	−.07	+.41	+.62	−.53	.00	−.01	−.46	+.49
Fourth	−.08	−.33	+.64	+.54	−.55	+.23	−.29	−.07	+.32	+.64	−.55	+.08	+.04	−.38	+.46

(a) In this computation the sums of the differences are assumed to be approximately zero. The error resulting from this assumption is small and decreases as the order of difference increases. For the second and higher differences the error resulting is not over .01. For first differences the error will not be over .01 except in those cases where the secular trend differs widely from a straight line, e. g., the compound interest law with a high rate of increase or decrease. In such cases the error may be as much as .04 or .05. However, in these cases the exact formula was used. The subscript of r indicates that the series designated by the letter first in alphabetical order precedes (−) or lags behind (+) the other series.

Digitized by Google

random we should expect the first differences to have 50 per cent. steadiness in sign and 50 per cent. alternation, on the average. We get exactly 50 per cent. as the average for the series listed. For second and higher differences the average percentages of alternation are: 64, 70, 75, 79, 82. A marked tendency to alternation in sign is revealed. The cumulative effect of this tendency to alternate in sign as higher differences are taken is also revealed in Table IV. Where the series *A*, *B*, *C*, *D*, and *E* are correlated with themselves (*AA*, *BB*, *CC*, *DD*, *EE*) there is a numerically increasing but negative coefficient for a lag of one item. Where all possible combinations of the five series are taken the coefficients alternate in sign or show a strong tendency to do so with the higher differences.

TABLE V.

COEFFICIENTS OF CORRELATION BETWEEN TWO RANDOM SERIES OF 35 ITEMS EACH AND BETWEEN THEIR MULTIPLE DIFFERENCES, FIRST TO EIGHT, CONCURRENT AND A LAG OF ONE AND OF TWO ITEMS IN EITHER DIRECTION.

Items Paired.	Coefficients of Correlation.				
	r_{-2}.	r_{-1}.	r_0.	r_{+1}.	r_{+2}.
Originals	+.32	+.08	−.11	−.42	−.05
First Differences	+.32	−.03	+.04	−.29	+.01
Second Differences	+.30	−.15	+.12	−.21	+.02
Third Differences	+.28	−.24	+.21	−.19	+.04
Fourth Differences	+.28	−.28	+.23	−.20	+.07
Fifth Differences	+.31	−.32	+.27	−.18	+.10
Sixth Differences	+.31	−.34	+.29	−.21	+.13
Seventh Differences	+.31	−.36	+.32	−.26	+.18
Eighth Differences	+.30	−.36	+.34	−.30	+.22

This theory of the tendency of signs of terms of higher differences to alternate and, therefore, to affect the coefficients of correlation was tested by applying the method of variate differences to two random series of 35 items each. The method of selection of the numbers was this: the pages of a table of six-place logarithms were turned at random, the tip of a pointer was placed at random on the page and the two digits at the right of the logarithm indicated by the pointer were taken as the items of the series. The coefficients of correlation between the two random series of 35 items and between their multiple differences, first to eighth, concurrent

Digitized by Google

and for one and two items lag in each direction are given in Table V. For the first and higher differences there is a persistent alternation in sign as we take a lag in either direction.* The coefficients alternate in sign, of course, because, first, the two series correlated alternate in sign, second, the terms alternating in sign become the dominating ones when the products of corresponding items are taken and third, lagging either series in either direction will bring a different set of signs into correspondence. The term Σxy tends to alternate in sign and hence r does.

III.

Does this phenomenon of the alternation in sign of the coefficients, left to right in the tables, have any bearing on the steadiness or unsteadiness of the coefficients, in the tabular columns, based on successive differences but with the same lag throughout? This question will now be considered.

The series of nth differences is derived from the series of $(n-1)$th differences by the same process that the series of first differences is derived from the original figures. Therefore, the expression for the coefficient of correlation for nth difference is the same function of the $(n-1)$th differences that the first differences is of the original series.

Let $r_L^{(m)}$ represent the coefficient of correlation between the mth differences of the series $x_o, x_1, \ldots x_{n-1}$ and $y_o, y_1, \ldots y_{n-1}$ where the subscript L denotes the lag of the x series. When $L=1$ we have the pairs $x_1y_o, x_2y_1 \ldots x_{n-1}\, y_{n-2}$; when $L=-1$ we have the pairs $x_oy_1, x_1y_2 \ldots x_{n-2}\, y_{n-1}$, etc. The formula for the coefficient of correlation between concurrent items of the original series is usually written in the form

$$r_o = \frac{\sum_{K=o}^{K=n-1} (x_K-\bar{x})(y_K-\bar{y})}{\sqrt{\sum_{K=o}^{K=n-1} (x_K-\bar{x})^2 \sum_{K=o}^{K=n-1} (y_K-\bar{y})^2}} \qquad (1)$$

when $\bar{x}$ and $\bar{y}$ are arithmetic averages of the respective series.

* None of the coefficients found signify appreciable correlation between the series. In but one case is the coefficient more than three times its probable error and in the majority of cases the probable error is approximately the same as the coefficient.

Digitized by Google

The function r_o may be written

$$r_o = \frac{n\sum_{o}^{n-1} x_K y_K - \sum_{o}^{n-1} x_K \sum_{o}^{n-1} y_K}{\sqrt{[n\sum_{o}^{n-1} x_K^2 - (\sum_{o}^{n-1} x_K)^2][n\sum_{o}^{n-1} y_K^2 - (\sum_{o}^{n-1} y_K)^2]}} \quad (2)$$

The coefficient of correlation r'_o between concurrent first differences x_1-x_o, y_1-y_o; x_2-x_1, y_2-y_1; . . . $x_{n-1}-x_{n-2}$, y_{n-1} y_{n-2} is, noticing that $\sum_{1}^{n-1}(x_K-x_{K-1}) = x_{n-1}-x_o$ and $\sum_{1}^{n-1}(y_K-y_{K-1}) = y_{n-1}-y_o$, *in terms of the original items.*

$$r'_o = \frac{(n-1)\sum_{1}^{n-1}(x_K-x_{K-1})(y_K-y_{K-1}) - (x_{n-1}-x_o)(y_{n-1}-y_o)}{\left\{\sqrt{[(n-1)\sum_{1}^{n-1}(x_K-x_{K-1})^2 - (x_{n-1}-x_o)^2]}\cdot \sqrt{[(n-1)\sum_{1}^{n-1}(y_K-y_{K-1})^2 - (y_{n-1}-y_o)^2]}\right\}} \quad (3)$$

Assuming $x_{n-1}-x_o$ and $y_{n-1}-y_o$ to be negligible in comparison with other terms of the function (called assumption *a*) we will discard the former. The function becomes

$$r'_o = \frac{2\sum_{o}^{n-1} x_K y_K - (x_o y_o + x_{n-1}y_{n-1}) - (\sum_{1}^{n-1} x_{K-1}y_K + \sum_{1}^{n-1} x_K y_{K-1})}{\left\{\sqrt{[2\sum_{o}^{n-1} x_K^2 - (x_o^2+x_{n-1}^2) - 2\sum_{1}^{n-1} x_K x_{K-1}]}\cdot \sqrt{[2\sum_{o}^{n-1} y_K - (y_o^2+y_{n-1}^2) - 2\sum_{1}^{n-1} y_K y_{K-1}]}\right\}} \quad (4)$$

Assuming that the original items of each series are random in order, first with respect to the adjacent items of the same series and, second, with respect to the items adjacent to the concurrent item of the other series, that is assuming $\sum_{1}^{n-1} x_{K-1}y_K$, $\sum_{1}^{n-1} x_K y_{K-1}$, $\sum_{1}^{n-1} x_K x_{K-1}$, and $\sum_{1}^{n-1} y_K y_{K-1}$ all equal zero (called assumption *b*) we have

$$r'_o = \frac{2\sum_{o}^{n-1} x_K y_K - (x_o y_o + x_{n-1}y_{n-1})}{\sqrt{[2\sum_{o}^{n-1} x_K^2 - (x_o^2+x_{n-1}^2)][2\sum_{o}^{n-1} y_K^2 - (y_o^2+y_{n-1}^2)]}} \quad (5)$$

Digitized by Google

Assuming $x_o y_o + x_{n-1} y_{n-1}$, $x_o^2 + x_{n-1}^2$ and $y_o^2 + y_{n-1}^2$ to be negligible in comparison with other terms in the function (called assumption *c*) we have, discarding the three terms named,

$$r'_o = \frac{\sum_o^{n-1} x_K y_K}{\sqrt{\sum_o^{n-1} x_K^2 \cdot \sum_o^{n-1} y_K^2}} = r_o, \text{ approximately, if } \bar{x} = \bar{y} = o$$

That is, if at any time assumptions *a*, *b*, and *c* hold true for any series of multiple differences, the coefficient of correlation for their first differences will equal the coefficient for the items from which the first differences were derived. If the same assumptions (*a*, *b*, and *c*) hold true for the series of first differences then $r''_o = r'_o$. Further if the assumptions hold true repeatedly for successive differences from and after the *p*th, we have $(m > p)$

$$r^{(m)} = r^{(m-1)} = \quad . \quad . \quad . \quad = r^{(p+1)} = r^{(p)}$$

It is obvious then, that the coefficients of correlation for successive differences will remain stable if assumptions *a*, *b*, and *c* hold true for successive differences. The condition (assumption that *a*, *b*, and *c* hold true repeatedly) is sufficient to produce stability, but is it necessary?

Consider form (4) of the function r'_o. If the two terms, $\sum_1^{n-1} x_K x_{K-1}$ and $\sum_1^{n-1} y_K y_{K-1}$, appearing in the denominator are negative in sign, if the expression $(\sum_1^{n-1} x_{K-1} y_K + \sum_1^{n-1} x_K y_{K-1})$ is opposite in sign to $\sum_1^{n-1} x_K y_K$, and if the terms are of appropriate size as well as sign, if these assumptions hold true, dropping $\sum_1^{n-1} x_K x_{K-1}$, $\sum_1^{n-1} y_K y_{K-1}$, and $(\sum_1^{n-1} x_{K-1} y_K + \sum_1^{n-1} x_K y_{K-1})$ will not affect the value of r'_o. Also r'_o will approximately equal r_o. In other words the fulfilling of the assumptions named will result in stable coefficients for successive differences. It is my contention that for a moderate number of items ($n = 35$ or 40) the conditions here specified are apt to occur and be the cause of any stability of the coefficients of correlation between multiple differences.

Digitized by Google

TABLE VI.

COEFFICIENTS OF CORRELATION BETWEEN THE BUSINESS BAROMETER AND THE CLEARING INDEX FOR THE UNITED STATES, FIRST TO SIXTH DIFFERENCES, CONCURRENT AND LAG AT ONE AND TWO ITEMS IN EITHER DIRECTION, 1879–1913.

Difference.	Coefficients of Correlation.			
	r_{-1}.	r_0.	r_{+1}.	r_{+2}.
First	−.33	+.68	+.39	−.22
Second	−.65	+.59	+.17	−.35
Third	−.71	+.57	+.02	−.26
Fourth	−.72	+.57	−.07	−.19
Fifth	−.71	+.57	−.14	−.06
Sixth	−.71	+.58	−.25	+.11

The results of applying the variate difference correlation method to the business barometer and the clearing index (not corrected for prices) are tabulated in Table VI. The first to sixth differences were used and the items were lagged in both directions. The stability of the coefficients for concurrent items second to sixth differences (r_o^{ii}, r_o^{iii}, r_o^{iv}, r_o^{v}, r_o^{vi}) and the instability of the coefficients for one item lag in the business barometer (r_{+1}^{iii} to r_{+1}^{vi}) are noticeable. Let us investigate the stability of r_o^{ii} and r_o^{iv} and the instability of r_{+1}^{iii} and r_{+1}^{v}. Using from (4) we have these values:

Term.		r_o^{iii}	r_o^{iv}	r_{+1}^{iii}	r_{+1}^{v}
$2\Sigma x_K y_K$	=	+3134	+2718	+ 888	− 950
$(x_o y_o + x_{n-1} y_{n-1})$	=	− 10	− 10	− 9	+ 12
$(\Sigma x_{K-1} y_K + \Sigma x_K y_{K-1})$	=	−1293	− 1618	+ 681	+ 2506
$2\Sigma x_K^2$	=	+7918	+ 2126	+7676	+ 6128
$(x_o^2 + x_{n-1}^2)$	=	+ 122	+ 58	+ 122	+ 37
$2\Sigma x_K x_{K-1}$	=	−3132	− 1276	−2890	− 4254
$2\Sigma y_K^2$	=	+3596	+10536	+3498	+32814
$(y_o^2 + y_{n-1}^2)$	=	+ 50	+ 250	5	+ 1360
$2\Sigma y_K y_{K-1}$	=	−1722	− 6298	−1694	−21556

Digitized by Google

Computing r_o''', r_o'' and r_{+1}''' and r_{+1}'' in two ways, by form (4) and form (5), we have the following results:

	Form (4).	Form (5).
r_o'''	+.59	+.59
r_o''	+.58	+.59
r_{+1}'''	+.03	+.17
r_{+1}''	−.15	−.07

The stability of r_o''' and r_o'' is explainable by the balancing effect of items $(\Sigma x_{K-1}y_K+\Sigma x_K y_{K-1})$, $\Sigma x_K x_{K-1}$ and $\Sigma y_K y_{K-1}$, the instability of r_{+1}''' and r_{+1}'' is explainable by the lack of balance between those items as they appear in the function. The items named are not approximately zero, in the case of r_o''' and r_o'', as "Student" assumes, but numerically as large as $\Sigma x_K y_K$.

TABLE VII.

COEFFICIENTS OF CORRELATION BETWEEN CONCURRENT ITEMS OF BUSINESS BAROMETER AND PRICES, RAILROAD GROSS EARNINGS, CORRECTED CLEARING INDEX AND SURPLUS RESERVES OF NEW YORK BANKS COMPUTED (I) BY EXACT FORMULA, (II) BY DISCARDING $\Sigma(x_K-x_{K-1})$ AND $\Sigma(y_K-y_{K-1})$ AND (III) BY DISCARDING $\Sigma x_{K-1}y_K-\Sigma x_K y_{K-1}$, $\Sigma x_K x_{K-1}$ AND $\Sigma y_K y_{K-1}$ FROM FORMULA.

	Prices.			R. R. Gross Ear.			Corr. Cl. Index.			Surplus Reserves.		
	I	II(a)	III(b)	I	II(a)	III(b)	I	II(a)	III(b)	I	II(a)	III(b)
Cycles	+.95	+.94	..	+.94	+.94	..	+.77	+.76	..	−.60	−.60	..
1st D.	+.78	+.78	+.94	+.80	+.80	+.94	+.32	+.32	+.76	−.72	−.71	−.60
2nd D.	+.78	+.78	+.79	+.74	+.74	+.81	+.23	+.23	+.34	−.74	−.74	−.72
3rd D.	+.78	+.78	+.79	+.74	+.74	+.77	+.21	+.21	+.23	−.75	−.75	−.74
4th D.	+.76	+.76	+.78	+.76	+.76	+.76	+.21	+.21	+.22	−.77	−.77	−.75
5th D.	+.74	+.74	+.76	+.75	+.75	+.76	+.24	+.24	+.22	−.78	−.78	−.77
6th D.	+.71	+.71	+.73	+.74	+.74	+.74	+.28	+.28	+.26	−.81	−.81	−.79

(a) The results in columns II coincide with those in columns I because the sums of the variate differences are approximately zero. See Table VIII.

(b) The results in columns III are equal to those of columns II of a lower order of difference. Approximate steadiness is reached not because the terms discarded are approximately zero but because they "balance." Evidence that the terms discarded are large is given by the coefficients of correlation for a lag in Table IV.

Digitized by Google

Table VII gives the results of applying, (I) the exact formula, (II) the formula obtained by assuming $\Sigma(x_K - x_{K-1}) = 0$, $\Sigma(y_K - y_{K-1}) = 0$ and (III) the formula obtained by assuming that the terms $(\Sigma x_{K-1}y_K - \Sigma x_K y_{K-1})$, $\Sigma x_K x_{K-1}$ and $\Sigma y_K y_{K-1}$ balance and therefore may be discarded, of applying these formulae to the business barometer and each of the series, wholesale prices, railroad gross earnings, corrected clearing index, and surplus reserves of New York associated banks. Form II gives coefficients practically identical with those resulting from the exact Form I showing that the assumption involved can safely be made. Table VIII illustrates the negligibility of the algebraic sums by comparing the algebraic sums of the second differences of various series with the absolute sums of the same items. Form III of Table VII gives coefficients which are equal to those for preceding differences as would be expected.

TABLE VIII.

ALGEBRAIC SUMS AND ABSOLUTE SUMS OF THE SECOND DIFFERENCES OF SIX SERIES WITH RATIO OF FORMER TO LATTER.

Second Differences of:	Algebraic Sum.	Absolute Sum.	Ratio.
Business barometer	0	150	.00
Wholesale prices	−11	157	−.07
Railroad gross earnings	+22	312	+.07
Clearing index	−7	277	−.03
Corrected clearing index	−6	456	−.01
Surplus reserves of New York banks	+2	568	.00

Consideration of the effects that assumptions *a*, *b*, and *c* have upon the value of the coefficients for multiple differences, leads to the conclusion that assumptions *a* and *c* are in accordance with the facts and that making them will not, for series of 35 or more items, in general affect the coefficient of correlation by more than .01. Assumption *b*, however, *does not* usually hold true. Nevertheless, stability is frequently secured by a balance between the terms appearing in numerator and denominator. Instability of the coefficients is explainable by lack of such balance. The fulfilling of assumption *c* is sufficient but not necessary to secure stability. Moreover, stability is usually secured when assumption *c* does not hold

Digitized by Google

true. In view of this conclusion, what significance has the variate difference correlation method for short economic time series? This question will now be considered.

IV.

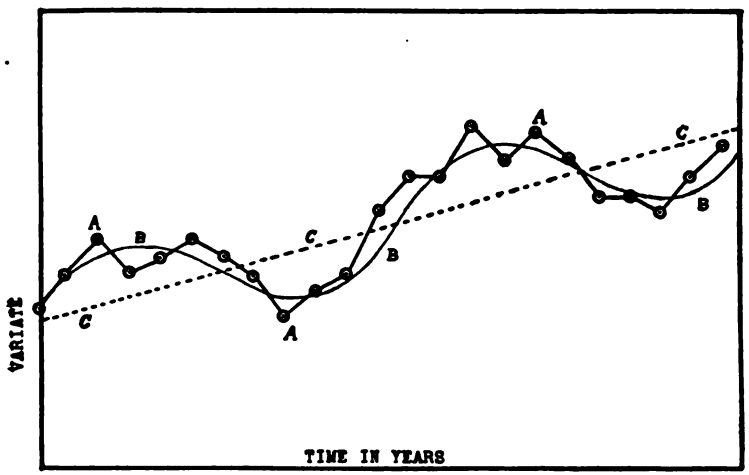
The items of annual time series of economic data may be conceived to be constituted of the following elements or component parts:

First, the *secular trend* or growth element due to the increase of population and development of industry;

Second, *cyclical fluctuations*, extending over a number of years and having a greater or less degree of periodicity, due to the alternating periods of business prosperity and depression;

Third, *irregular fluctuations* from year to year due to the influence of accidental or, at any rate, unpredictable events such as inventions, striking changes in fashion, or war.

FIGURE 1



A—REPRESENTS ECONOMIC DATA.
B—REPRESENTS CYCLICAL FLUCTUATION.
C—REPRESENTS SECULAR TREND.

Digitized by Google

An idealistic representation of a time series of economic data containing the three elements, secular trend, cyclical, and irregular fluctuation is given in Figure 1. For particular data the secular trend may, of course, be other than a straight line and the cyclical fluctuations other than the simple sine curve.

The concept of secular trend or normal growth element for which I contend is that of an element which increases or decreases regularly according to some principle. That principle may be linear, *i. e.*, the addition of a constant amount in each time interval as assumed in Figure 1, or it may be the compound interest law, the addition of an equal percentage in each time interval, or it may be a second degree parabola, or some other law. In any case, however, the mathematical function assumed to represent the secular trend should not "fit" the cyclical or irregular fluctuations of the data. Now the process of taking multiple differences is equivalent, on the assumption that the series is an algebraic function of time, to reducing the degree of the function of t by one for each difference taken. Thus, if the secular trend be linear, if a straight line should be fitted to the data and if the deviations of the original series from the corresponding ordinates of the straight line are taken (the cycles), then the coefficient of correlation between the first differences of the original items and that between the first differences of the cycles are identical. This theorem may be proven as follows:

Let $x_o, x_1, x_2 \ . \ . \ . \ x_{n-1}$ be the original series of n items and let $x=mt+b$ be the line of secular trend. The first differences are of the form x_K-x_{K-1}. The cycles are of the form $x_K-(mt_K+b)$. The first differences of the cycles are of the form $x_K-x_{K-1}-m(t_K-t_{K-1})$. Since t_K-t_{K-1} is a constant for all values of K (being the time interval for which items are taken) the first differences of the cycles differ from the first differences of the original series by a constant. Hence the coefficients of correlation obtained from them are identical.

The coefficient of correlation between second differences is identical with that found between the corresponding deviations of items from the smoothed curve obtained by taking

Digitized by Google

three-year averages.* This theorem may be proven as follows:

Let x_o, x_1, x_2 . . . x_{n-1} be the original series. The second differences are of the form $x_{K+1}-2x_K+x_{K-1}$. The deviation of any items from the three-year average is $x_K-\frac{x_{K+1}+x_K+x_{K-1}}{3}$ or $\frac{-x_{K+1}+2x_K-x_{K-1}}{3}$. The form just given is obtainable from the form for second differences by multiplying by $-\frac{1}{3}$. Therefore the coefficients of correlation between second differences are the same as those between deviations from three-year averages.†

* Moore has correlated deviations from three-year averages of crop yield per acre (weighted index of nine leading crops) and production of pig-iron. He does not appear to realise that his coefficients are the same as those for second differences. The coefficient for crops and pig-iron production with a one-year lag of the latter (r''_{+1}) for the period 1870-1911 is .264; for crops and general prices for the same period the coefficient (r''_o) is .303. See Moore's *Economic Cycles*, pp. 107, 118.

† (Theorem for form of nth difference)

Theorem: The nth difference of the terms of a series

$$x_1, x_2, x_3 \quad . \quad . \quad . \quad x_n$$

is of the form

$$x_K^{(m)}=[C_m x_{K+m}-C_{m-1}x_{K+m-1}+C_{m-2}x_{K+m-2}- \quad . \quad . \quad . \quad +(-1)^m C_o x_K]$$

where C_i indicates the number of combinations of m things taken i at a time.

The first differences of the series

x_1, x_2 . . . x_n are x_2-x_1, x_3-x_2 . . . $x_{K+1}-x_K$ . . . x_n-x_{n-1}.

The second differences are

$$x_3-2x_2+x_1 \quad . \quad . \quad . \quad x_{K+1}-2x_K+x_{K-1} \quad . \quad . \quad . \quad x_n-2x_{n-1}+x_{n-2}.$$

Assume the mth difference to be of the form

$$x_{K+m}-mx_{K+m-1}+\frac{m(m-1)}{\underline{|2}}x_{K+m-2}- \quad . \quad . \quad . \quad +(-1)^m x_K \text{ and}$$

$$x_{K+m+1}-mx_{K+m}+\frac{m(m-1)}{\underline{|2}}x_{K+m-1}- \quad . \quad . \quad . \quad +(-1)_m x_{K+1}$$

Taking the difference we have

$$x_{K+m+1}-(m+1)x_{K+m}+\frac{m(m-1)-2m}{\underline{|2}}x_{K+m-1} \quad . \quad . \quad .$$

$$. \quad . \quad . \quad \pm\left[\frac{\underline{|m}}{\underline{|i}\,\underline{|m-i}}+\frac{\underline{|m}}{\underline{|i-1}\,\underline{|m-i+1}}\right]x_{K+m-i+1} \quad . \quad . \quad .$$

Consider the coefficient of the $(i+1)$th term:

$$\frac{\underline{|m}}{\underline{|i}\,\underline{|m-i}}+\frac{\underline{|m}}{\underline{|i-1}\,\underline{|m-i+1}}, \text{ which reduces to}$$

$$\underline{|m}\left[\frac{(m-i+1)+i}{\underline{|i}\,\underline{|m-i+1}}\right]=\frac{\underline{|m+1}}{\underline{|i}\,\underline{|m+1-i}}=C_i^{m+1}$$

Since we have shown that the 2nd differences are expressions with the binomial coefficients corresponding to the order of difference and since, *assuming* that the mth difference appears with the binomial coeffi-

Digitized by Google

These theoretical considerations taken in connection with applications and tests of the variate difference method lead me to the following conclusions concerning the meaning of the coefficients of correlation for the raw figures and their multiple differences:

In general, significant coefficients of correlation for the raw figures of two time series indicate the similarity of the growth elements of the two series, *if large growth elements exist.* The existence or non-existence of such elements is readily determined graphically or by fitting a simple function to the data.

Significant coefficients of correlation for first differences indicate that the cyclical fluctuations synchronize, *if there be cyclical fluctuations.* Evidence of such cycles may be secured by plotting the deviations from the assumed secular trend.

Significant coefficients of correlation for second and in some cases higher differences indicate, in general, that the irregular fluctuations synchronize. Coefficients for higher differences of short series contain a large spurious element which increases with the order of difference. This element is due to the tendency of the items to alternate in sign.

These conclusions assume that the magnitude of the elements due to secular trend, business cycles, and irregular events vary in the order in which these elements are named. In case we are dealing with data having marked cycles (of the same variety) and are interested in the correlation of the cycles the coefficients for the cycles and first differences constitute the proper basis for judgment rather than coefficients for higher differences. Stability of coefficients for higher differences, in such a case, probably means that the influence of all the fluctuations except the irregular ones, *i. e.*, the oscillations, has been eliminated by the variate difference process.

cients corresponding to the terms of the expansion of $(1+1)^m$, we have proven that the $(m+1)$th difference has as coefficients the terms of the expansion of $(1+1)^{m+1}$ then, by mathematical induction, the proposition stated is true.

Professor A. A. Young has called my attention to the fact that the foregoing theorem is demonstrated in the *Institute of Actuaries' Text Book*, 2nd ed., Part II, p. 427.

Digitized by Google

TABLE IX.

LONDON CLEARINGS (COLUMN 1) WITH THE 9-YEAR MOVING AVERAGE (2) AND DEVIATIONS FROM MOVING AVERAGE (3) STRAIGHT LINE (4), PARABOLA (5) AND COMPOUND INTEREST LAW (6); SAUERBECK'S PRICES (7) WITH THE 9-YEAR MOVING AVERAGE (8) AND DEVIATIONS FROM MOVING AVERAGE (9), STRAIGHT LINE (10) AND PARABOLA (11); 1868–1913. ALSO DEVIATIONS OF LONDON CLEARINGS FROM THE TWO STRAIGHT LINES FITTED TO DATA FOR 1868–1896 AND FOR 1897–1913 (12) AND DEVIATIONS OF SAUERBECK'S PRICES FROM TWO STRAIGHT LINES SIMILARLY FOUND (13).

Year.	1.(a)	2.	3.	4.	5.	6.	7.(b)	8.	9.	10.	11.	12.	13.
1868	34	....	....	+ 6	−18	− 3	99	....	....	+ 8	−11	−10	− 7
1869	36	....	....	+ 5	−15	− 2	98	....	....	+ 8	−10	− 9	− 6
1870	39	....	....	+ 6	−11	− 1	96	....	....	+ 6	− 9	− 7	− 6
1871	48	....	....	+13	− 2	+ 7	100	....	....	+11	− 2	+ 1	− 1
1872	59	49	+10	+21	+10	+17	109	101	+ 8	+21	+10	+11	+10
1873	61	51	+10	+21	+12	+17	111	100	+11	+23	+14	+12	+13
1874	59	53	+ 6	+17	+10	+15	102	99	+ 3	+15	+ 7	+ 9	+ 6
1875	57	54	+ 3	+12	+ 9	+11	96	97	− 1	+10	+ 4	+ 6	+ 2
1876	50	55	− 5	+ 3	+ 1	+ 2	95	96	− 1	+ 9	+ 5	− 2	+ 2
1877	50	55	− 5	+ 1	+ 1	+ 1	94	93	+ 1	+ 9	+ 6	− 3	+ 3
1878	50	56	− 6	− 1	+ 1	− 1	87	90	− 3	+ 3	+ 1	− 4	− 2
1879	49	56	− 7	− 5	− 1	− 3	83	88	− 5	− 1	− 2	− 7	− 5
1880	58	56	+ 2	+ 2	+ 8	+ 4	88	86	+ 2	+ 5	+ 5	+ 1	+ 2
1881	64	56	+ 8	+ 6	+14	+ 8	85	83	+ 2	+ 2	+ 4	+ 6	+ 1
1882	62	57	+ 5	+ 1	+10	+ 5	84	81	+ 3	+ 2	+ 4	+ 3	+ 1
1883	59	58	+ 1	− 4	+ 6	0	82	79	+ 3	+ 1	+ 4	− 1	+ 1
1884	58	61	− 3	− 7	+ 3	− 3	76	77	− 1	− 5	− 1	− 3	− 3
1885	55	62	− 7	−13	− 1	− 8	72	75	− 3	− 8	− 4	− 7	− 6
1886	59	64	− 5	−11	+ 2	− 6	69	74	− 5	−10	− 5	− 4	− 7
1887	61	65	− 4	−11	+ 2	− 6	68	73	− 5	−11	− 5	− 3	− 7
1888	69	65	+ 4	− 6	+ 8	+ 1	70	71	− 1	− 8	− 2	+ 4	− 3
1889	76	66	+10	− 1	+13	+ 5	72	70	+ 2	− 6	+ 1	+10	+ 1
1890	78	67	+11	− 1	+13	+ 5	72	69	+ 3	− 5	+ 1	+10	+ 2
1891	69	69	0	−13	+ 2	− 7	72	68	+ 4	− 4	+ 2	0	+ 4
1892	65	71	− 6	−19	− 5	−13	68	68	0	− 8	− 2	− 5	+ 2
1893	65	71	− 6	−21	− 7	−15	68	67	+ 1	− 7	− 1	− 6	+ 3
1894	63	72	− 9	−25	−12	−19	63	66	− 3	−11	− 6	− 9	0
1895	76	73	+ 3	−15	− 2	− 9	62	65	− 3	−12	− 7	+ 3	+ 1
1896	76	76	0	−17	− 5	−12	61	66	− 5	−12	− 8	+ 2	+ 1
1897	75	79	− 4	−20	− 9	−16	62	66	− 4	−10	− 7	0	− 2
1898	81	83	− 2	−17	− 7	−12	64	66	− 2	− 8	− 5	+ 1	− 2
1899	92	87	+ 5	− 8	+ 1	− 5	68	67	+ 1	− 3	− 1	+ 7	+ 1
1900	90	91	− 1	−12	− 5	−10	75	68	+ 7	+ 4	+ 6	− 1	+ 7
1901	96	96	0	− 9	− 2	− 7	70	69	+ 1	0	+ 1	0	+ 1
1902	100	102	− 2	− 7	− 3	− 5	69	70	− 1	0	− 1	− 1	− 1
1903	101	107	− 6	− 8	− 6	− 8	69	72	− 3	0	− 1	− 6	− 2
1904	106	110	− 4	− 6	− 5	− 7	70	73	− 3	+ 2	− 1	− 6	− 2
1905	123	115	+ 8	+ 9	+ 8	+ 7	72	73	− 1	+ 5	0	+ 6	− 2
1906	127	121	+ 6	+11	+ 7	+ 8	77	74	+ 3	+10	+ 4	+ 5	+ 2
1907	127	126	+ 1	+ 9	+ 2	+ 4	80	75	+ 5	+14	+ 6	− 1	+ 4
1908	131	132	−11	0	− 8	− 6	73	76	− 3	+ 7	− 2	−12	− 4
1909	135	139	− 4	+12	+ 1	+ 5	74	78	− 4	+ 9	− 2	− 3	− 4
1910	147	....	....	+22	+ 8	+12	78	....	....	+14	+ 1	+ 3	− 1
1911	146	....	....	+18	+ 1	+ 7	80	....	....	+16	+ 1	− 3	− 1
1912	160	....	....	+30	+10	+16	85	....	....	+22	+ 5	+ 6	+ 3
1913	164	....	....	+32	+ 8	+17	85	....	....	+23	+ 3	+ 4	+ 2

(a) In £100,000,000. (b) Relative Indices.

Judgment concerning the correlation of cyclical fluctuations of two series must be preceded by elimination of the secular trend. The choice of a function to represent the secular trend, indeed the choice of the method of eliminating the trend,

Digitized by Google

whether by curve fitting or otherwise, these are questions fundamental to the process. I will test the effect that various suppositions concerning the secular trend have upon the correlation coefficients between the deviations from the various trends (cycles) resulting from the suppositions. The two series chosen for this test are London clearings and Sauerbeck's price indices given with other data in Table IX. These series are chosen, first, because the secular trends are dissimilar, second, because the trends differ most widely from the linear of any which could be found, and, third, because the variate difference correlation coefficients for these series are puzzling to the author of the variate difference correlation method. "Student" applied his method to London clearings per capita and Sauerbeck's prices and to marriage rate and wages, finding the following coefficients:

I. CLEARINGS AND PRICES.

Raw Figures.	1st D.	2nd D.	3rd D.	4th D.	5th D.	6th D.
−.33	+.51	+.30	+.07	+.11	+.05	..

II. MARRIAGE RATE AND WAGES.

−.52	+.67	+.58	+.52	+.55	+.58	+ 55

He says, "The difference between I and II is very marked, and would seem to indicate that the causal connection between index numbers and Bankers' clearing house rates is not altogether of the same kind as that between marriage rate and wages, though all four variables are commonly taken as indications of the short period trade wave."*

Figure 2 presents Sauerbeck's indices with linear and parabolic secular trends, the functions being fitted to the data by the method of moments. Figure 3 presents London bank clearings with linear, parabolic, and exponetial functions fitted to the data. Figure 4 presents both series with their respective nine-year moving averages, nine years being deter-

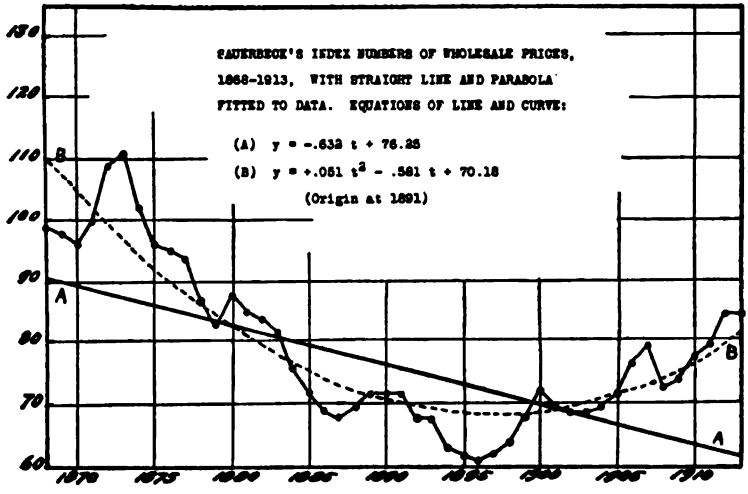
* *Biometrika*, April, 1914, p. 180.

Digitized by Google

mined by inspection as the length of the business wave or cycle. Figures 5, 6, 7, and 8 present the deviations, positive or negative, of the two series from the various secular trends. The last named figures all show a striking correspondence of the cyclical fluctuations of the two series. It will be noticed that fluctuations of clearings show a tendency to precede or forecast the fluctuations in prices.

Figures 5, 6, 7, and 8 throw some light on Babson's hypothesis that economic action and reaction are equal, *i. e.*, that consecutive areas above and below the line of normal growth should be equal for a correct normal line. It is true that the *sum* of areas above and below the lines are roughly equal; that, the method of curve fitting accomplishes. But what are *consecutive* areas, the long-time areas of Figures 6 and 8 or the short-time areas of Figures 5 and 7? It is obvious that we would get still more heterogeneous results, as regards positive and negative areas, if we should use series of various lengths, say 20 years or 70 years, or series of monthly or quarterly rather than annual data.

FIGURE 2.



Digitized by Google

FIGURE 3

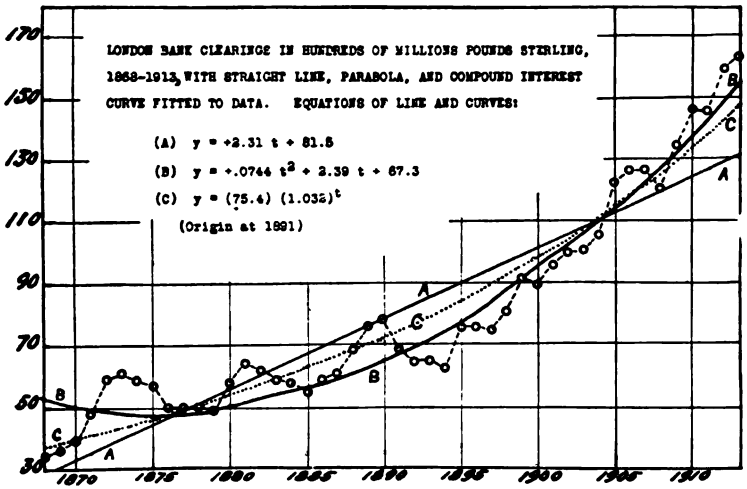
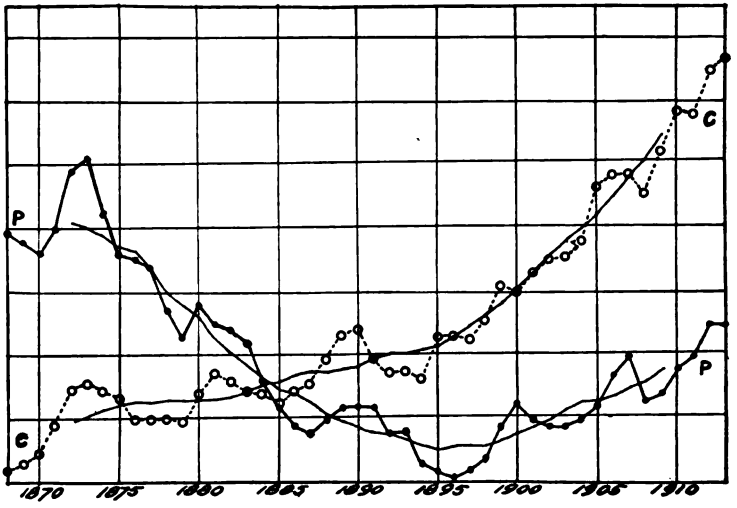


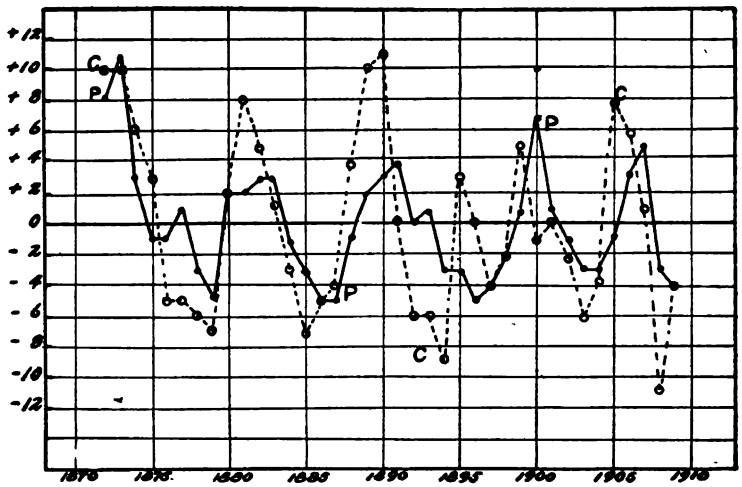
FIGURE 4



SAUERBECK'S PRICE INDICES (P) AND LONDON CLEARINGS (C), 1868-1913, WITH THEIR RESPECTIVE NINE-YEAR MOVING AVERAGES, 1872-1909.

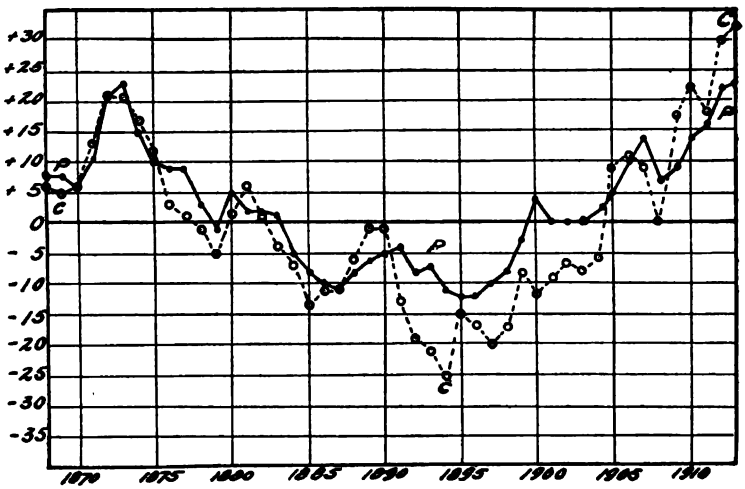
Digitized by Google

FIGURE 5



DEVIATIONS OF LONDON CLEARINGS (C) AND SAUERBECK'S PRICE INDICES (P) FROM THEIR RESPECTIVE NINE-YEAR MOVING AVERAGES AS SECULAR TRENDS, 1872–1909.

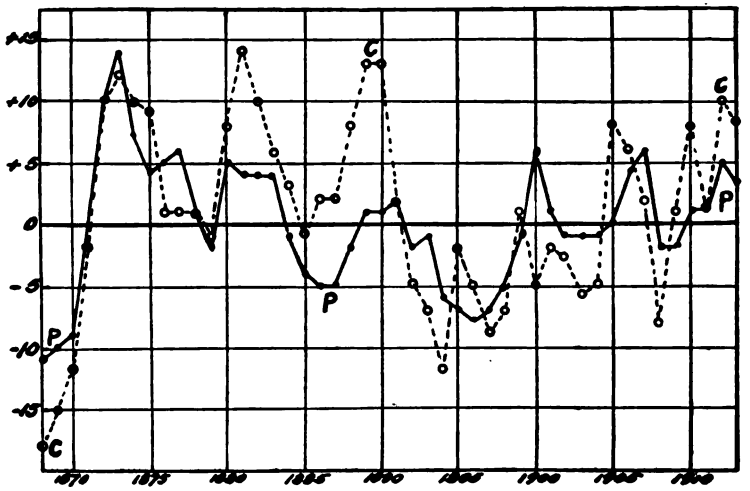
FIGURE 6



DEVIATIONS OF LONDON CLEARINGS (C) AND SAUERBECK'S PRICES (P) FROM THEIR RESPECTIVE LINEAR SECULAR TRENDS, 1868–1913.

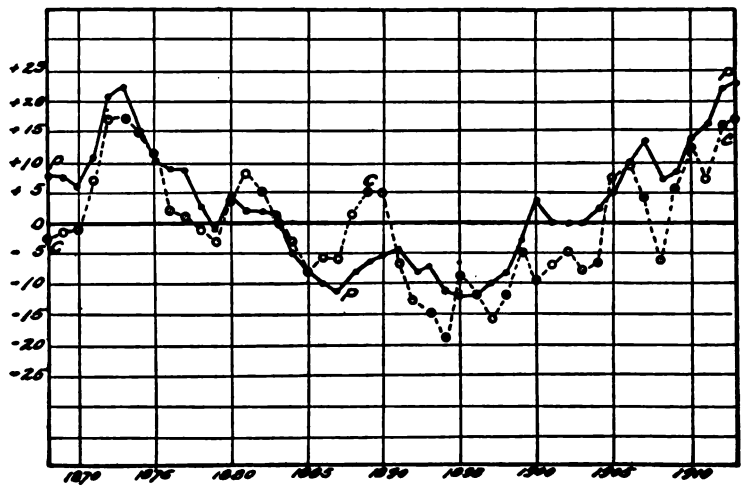
Digitized by Google

FIGURE 7



DEVIATIONS OF LONDON CLEARINGS (C) AND SAUERBECK'S PRICES (P) FROM THEIR RESPECTIVE PARABOLIC SECULAR TRENDS, 1868–1913.

FIGURE 8



DEVIATIONS OF LONDON CLEARINGS FROM TREND AS COMPOUND INTEREST CURVE (C) AND AT SAUERBECK'S PRICES FROM LINEAR TREND (P), 1868–1913.

Digitized by Google

TABLE X.

COEFFICIENTS OF CORRELATION BETWEEN SAUERBECK'S PRICE INDICES AND LONDON CLEARINGS, 1868–1913.

A. Raw Figures and their Differences.
B. Deviations from 9-year Moving Average and Differences.
C. Deviations from Straight Lines and Differences.
D. Deviations from Parabolas and Differences.
E. Deviations from Compound Interest Law for Clearings and Straight Line for Prices and Differences.

	r_{-2}.	r_{-1}.	r_0.	r_{+1}.	r_{+2}.
A.					
Raw Figures	....	−.37	−.31	−.28	....
First Differences	....	+.05	+.42	+.37	+.12
Second Differences	....	−.20	+.31	+.14	....
Third Differences	....	−.27	+.26	+.03	....
Fourth Differences	....	−.26	+.23	−.01	....
Fifth Differences	....	−.28	+.23	−.06	....
Sixth Differences	....	−.21	+.21	−.08	....
B.					
Deviations	−.27	+.19	+.64	+.69	+.36
First Differences	....	−.06	+.42	+.37	....
Second Differences	....	−.15	+.22	+.10	....
Third Differences	....	+.10	+.15	+.02	....
C.					
Deviations	+.75	+.85	+.92	+.90	+.80
First Differences	....	+.09	+.55	+.43	....
Second Differences	....	−.21	+.35	+.08	....
Third Differences	....	−.26	+.31	−.04	....
D.					
Deviations	+.20	+.49	+.73	+.65	+.33
First Differences	....	+.06	+.49	+.36	....
Second Differences	....	−.22	+.35	+.06	....
Third Differences	....	−.28	+.31	−.06	....
E.					
Deviations	+.67	+.75	+.82	+.76	+.59
First Differences	....	+.02	+.50	+.38	....
Second Differences	....	−.26	+.33	+.13	....
Third Differences	....	−.32	+.35	.00	....

The main question upon which we wish to get light is, however, the effect of the various methods of eliminating the secular trend upon the coefficients of correlation between corresponding deviations. Table X gives the coefficients of

Digitized by Google

correlation between Sauerbeck's indices and London clearings taking the raw figures and their differences, first to sixth, and also taking deviations from various secular trends and their differences, first to third. Coefficients are presented for concurrent items and for a lag in both directions, a lag of one for differences and a lag of one and of two for deviations or cycles.

The coefficients of correlation for the raw figures (−.37, −.31 and −.28) show that the secular trends of prices and clearings are in opposite directions. The coefficients for the first differences of the raw figures and of all the deviations indicate an appreciable positive correlation for concurrent items (r'_0) and for prices one-year lag (r'_{+1}), with the coefficient r'_0 larger. The coefficients for second and higher differences of the raw figures, and deviations as well, decrease as the order of difference increases; the coefficients for one-year lag of prices decreasing more rapidly than for concurrent items. These facts indicate that the maximum correlation of business cycles (including the irregular fluctuations) is for clearings preceding prices by less than half a year, say, four months. There is, however, an unknown element of spurious correlation between clearings and prices because the former are dependent upon prices as well as physical volume of goods exchanged and speculation. If this spurious element, due not to the method but to the nature of the data, were excluded, it is probable that the maximum correlation would be found for clearings preceding prices by more than six months, perhaps by a year.

The coefficients of correlation for the deviations all agree in locating the maximum, and therefore the lag of prices, at less than a year. The actual maximum found was for concurrent items, except for deviations from the nine-year averages which gives a maximum at one year lag of prices. Since our judgment is based upon the relative values of the coefficients for various degrees of lag, rather than upon their absolute values, the type of secular trend chosen does not appear to have great significance. Curve-fitting, however, does appear to be preferable to the taking of moving averages because, first, all the items may be used in determining the correlation and, second, the coefficients for deviations and first differences disagree in their location of the maximum when deviations from the mov-

Digitized by Google

ing average are taken but agree in all other cases. Of course it might be possible to use the deviations of all the terminal items from the moving average, if values of the latter were exterpolated; fitting a curve, graphically or otherwise, to the moving average is the obvious solution of this problem.

The results appearing in Table X make it clear that the coefficients for higher differences give little indication of the correspondence of the business cycles in the two series, which correspondence is clearly shown by the charts and the correlation coefficients for deviations from the secular trend and for first differences.

V.

If our interest were in the absolute degree of correlation between the cycles of two series for selected pairs of items, the nature of the curves used to represent the secular trends and the closeness of the fit to the data would be of primary importance. In case we are dealing with series in which the secular trends are non-linear, such as clearings and prices, but if, nevertheless, we use straight lines to represent the trends and correlate deviations therefrom, the coefficients resulting will undoubtedly contain a large spurious element, positive or negative. This is illustrated by the discrepancy between the coefficients, +.92 and +.73, obtained for deviations from straight lines and parabolas, respectively, of London clearings and Sauerbeck's prices (Table X). The former coefficient (+.92) undoubtedly contains a spurious element amounting to some twenty points. The spurious element is positive in this case, apparently, because of the downward long time trend from 1868 to 1896 and the upward trend from 1897 to 1913 which results in the pairing of large negative deviations for the period 1884–1899, when the deviations are from straight lines.

To test the effect of dividing the data of Sauerbeck's prices and London clearings into two homogeneous sub-periods, *i. e.* one of falling prices, 1868–1896, and one of rising prices, 1897–1913, two linear secular trends were found for each series and coefficients of correlation were computed for deviations from these trends and for their first differences. The lines

Digitized by Google

and their equations are given in Figure 9; the deviations appear in Figure 10; the coefficients are presented in Table XI. The maximum coefficients for the deviations are $r_0 = +.71$ and $r_{+1} = +.70$ for the periods 1868–1896 and 1897–1913, respectively. The maximum coefficients for corresponding first differences, $r'_0 = +.61$ and $r'_{+1} = +.47$, are consistent with those obtained for the entire period as shown by Table X. The coefficients for first differences, r'_0, between deviations from various trends (see A, B, C, D, and E of Table X) are +.42, +.42, +.55, +.49, and +.50.

TABLE XI.

COEFFICIENTS OF CORRELATION BETWEEN DEVIATIONS OF SAUERBECK'S PRICES AND OF LONDON CLEARINGS FROM THEIR RESPECTIVE LINEAR SECULAR TRENDS FOR THE TWO PERIODS 1868–1896 AND 1897–1913 TOGETHER WITH COEFFICIENTS FOR FIRST DIFFERENCES; VARIOUS DEGREES OF LAG OF PRICES (+) AND OF CLEARINGS (−).

Items Paired.	Coefficients of Correlation.				
	r_{-1}.	r_0.	r_{+1}.	r_{+2}.	r_{+3}.
			1868–1896.		
Deviations	+.38	+.71	+.63	+.34	−.07
First Differences	+.20	+.61	+.34	+.16	−.07
			1897–1913.		
Deviations	−.27	+.43	+.70	+.20	−.34
First Differences	−.40	+.27	+.47	+.04	−.40

Division of the data into two sections throws new light on the problem. Clearings and prices fluctuated concurrently during the first period, but prices lagged behind clearings by a year during the period 1896–1913. Perhaps increased speculation has changed the character of clearings during the second period. Whatever may be the cause, the fluctuations of English prices and clearings are shown to be related in the same fashion as are those for the United States during the same period (see AD, BD, and CD of Table IV).

Digitized by Google

TABLE XII.

COEFFICIENTS OF CORRELATION BETWEEN RELATIVE WHOLESALE PRICES AND PIG-IRON PRODUCTION OF THE UNITED STATES, DEVIATIONS AND FIRST DIFFERENCES AS FOLLOWS:

1879–1913.

A. Deviations from Linear Secular Trends and First Differences.

B. Deviations of Prices from Parabola and of Pig-iron Production from Compound Interest Curve, and Their First Differences.

1879–1896.

C. Deviations from Linear Secular Trends and Their First Differences.

1897–1913.

D. Deviations from Linear Secular Trends and Their First Differences.

1879–1896.

E. Deviations of Prices from Linear Secular Trend, and of Pig-iron Production from Compound Interest Curve (Computed for Data 1879–1913).

1897–1913.

F. Deviations of Prices from Linear Secular Trend, and of Pig-iron Production from Compound Interest Curve (Computed for Data 1879–1913).

1879–1913.

G. Deviations of Prices from the Two Linear Secular Trends (1879–1896 and 1897–1913) as a Continuous Series and of Pig-iron Production from Compound Interest Curve.

H. First Differences of Raw Items.

Prices concurrent (0), lag (+), or previous (−) to pig-iron production as indicated by subscript of r.

Symbol.	Period.	Items Paired.	Coefficients of Correlation.				
			r_{-1}.	r_0.	r_{+1}.	r_{+2}.	r_{+3}.
(A)	1879–1913	Deviations (a) First Differences	..... +.31	+.76 +.41	+.74 +.21	+.63	+.62 .....
(B)	1879–1913	Deviations First Differences	+.05 −.37	+.48 +.40	+.51 +.22	....	....
(C)	1879–1896	Deviations First Differences	....	+.48 +.23	+.61 +.55	+.41 .00	+.31 .....
(D)	1897–1913	Deviations First Differences	..... −.65	+.41 +.54	+.35 +.17	−.06	−.06 .....
(E)	1879–1896	Deviations	.....	+.54	+.69	+.49	+.20
(F)	1897–1913	Deviations	.....	+.55	+.48	+.10	+.11
(G)	1879–1913	Deviations	.....	+.45	+.44	+.19	+.10
(H)	1879–1913	First Differences	−.31	+.41	+.22	−.10	+.07

(a) Coefficient $r_{+4} = +.53$.

Digitized by Google

Table XII presents coefficients of correlation between deviations from various secular trends and their first differences of relative wholesale prices and pig-iron production of the United States for the period 1879–1913 and the sub-periods, 1879–1896 and 1897–1913. The various secular trends and their equations are given in Figures 11 and 12; the deviations appear in Figures 13 and 14; the data appear in Table XIII.

For the period 1879–1913 (see A, B, G, and H) the cycles of prices and pig-iron production are concurrent. For the period 1879–1896 (see C and E) the pig-iron cycles precede price cycles by a year. For the period 1897–1913 the cycles of the two series are strongly concurrent (see D and F). The coefficients of correlation for deviations from the linear trends, 1879–1896, $r_0 = +.48$ and $r_{+1} = +.61$ and $r_{+2} = +.41$, agree in locating the maximum at the same point as those for first differences, $r'_0 = +.23$ and $r'_{+1} = +.55$ and $r'_{+2} = .00$ (see C). The coefficients for deviations from the linear trends, 1897–1913, $r_0 = +.41$ and $r_{+1} = +.35$ are likewise supported by those for first differences, $r'_0 = +.54$ and $r'_{+1} = +.17$ (see D). Using deviations of prices from the two linear secular trends as *a continuous series* and of pig-iron production from the compound interest curve, 1879–1913, we have the coefficients $r_0 = +.45$ and $r_{+1} = +.44$ (see G). It is obvious that the coefficients for the whole period and the two sub-periods are consistent. *At present general prices and pig-iron production fluctuate concurrently.*

Digitized by Google

FIGURE 9.

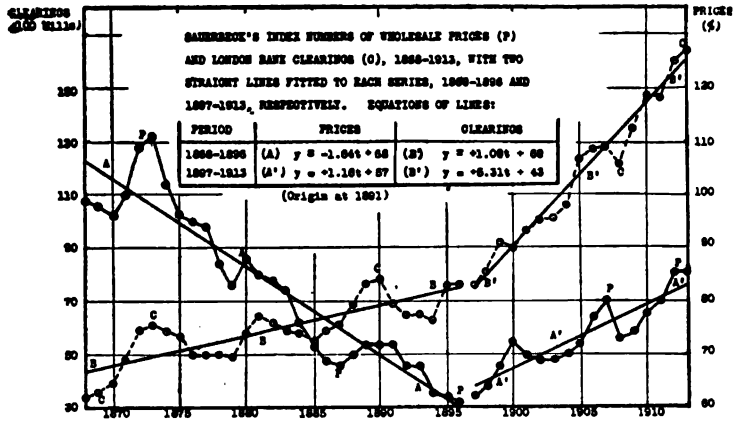
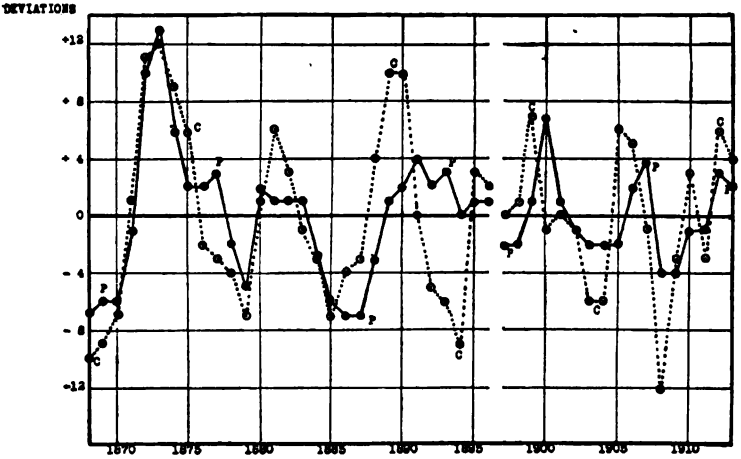


FIGURE 10.



DEVIATIONS OF LONDON CLEARINGS (C) AND SAUERBECK'S PRICE INDICES (P) FROM THEIR RESPECTIVE LINEAR SECULAR TRENDS FOR THE TWO PERIODS 1868-1896 AND 1897-1913.

Digitized by Google

FIGURE 11.

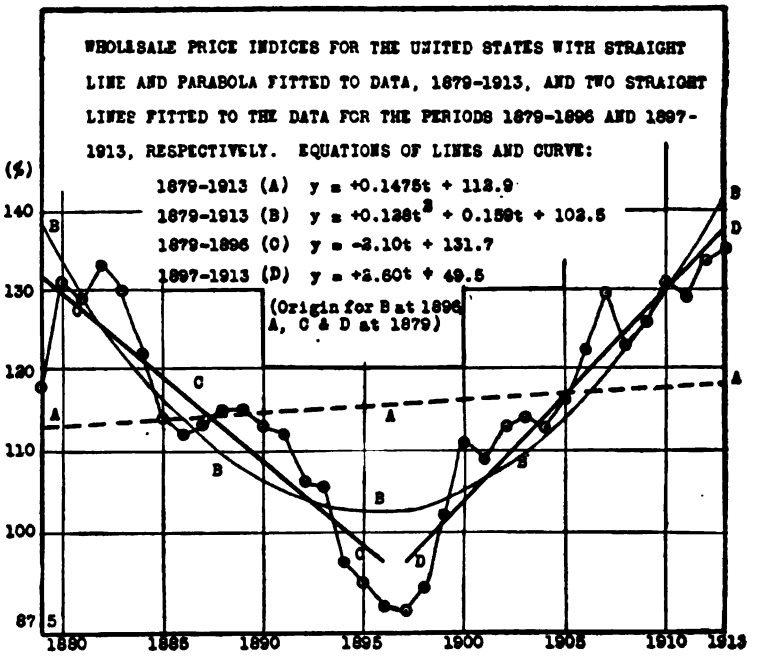
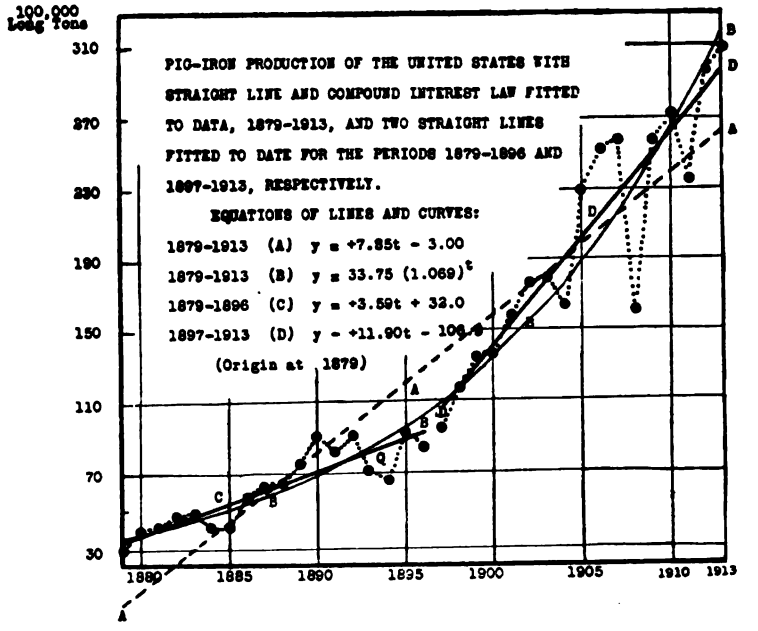
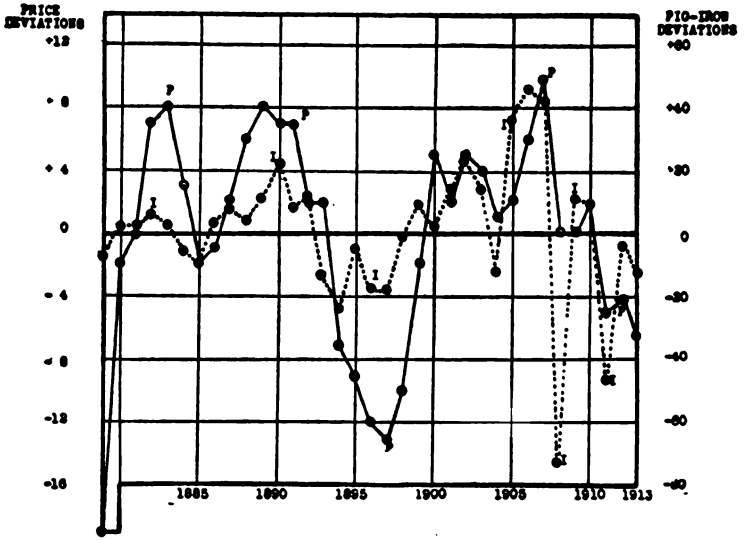


FIGURE 12.



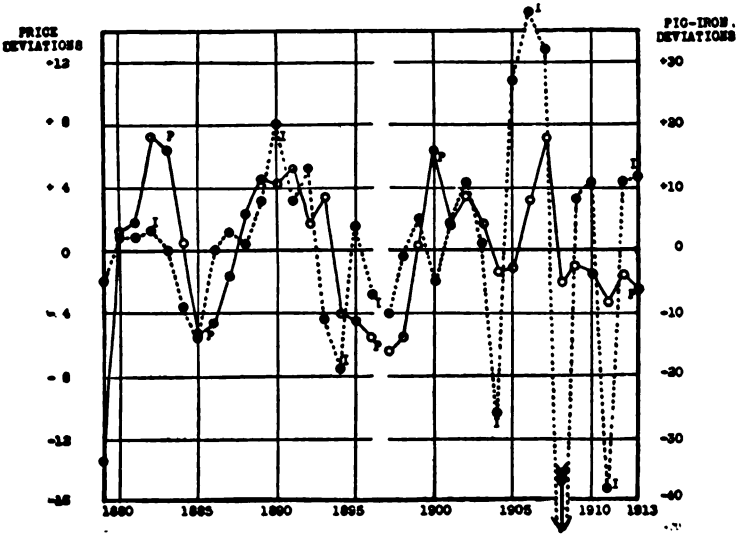
Digitized by Google

FIGURE 13.



DEVIATIONS OF UNITED STATES PRICE INDICES (P) FROM PARABOLA AND PIG-IRON PRODUCTION (I) FROM COMPOUND INTEREST CURVE, 1879-1913.

FIGURE 14.



DEVIATIONS OF UNITED STATES PRICE INDICES (P) AND PIG-IRON PRODUCTION (I) FROM THEIR RESPECTIVE LINEAR SECULAR TRENDS FOR THE TWO PERIODS, 1879-1896 AND 1897-1913.

Digitized by Google

TABLE XIII.

WHOLESALE PRICE INDICES FOR THE UNITED STATES (COLUMN 1) WITH THE DEVIATIONS FROM STRAIGHT LINE (2) AND PARABOLA (3) 1879-1913, AND FROM TWO STRAIGHT LINES, 1879-1896 AND 1897-1913 (4) AS SECULAR TRENDS. ALSO PIG-IRON PRODUCTION IN THE UNITED STATES (5) WITH THE DEVIATIONS FROM STRAIGHT LINE (6) AND FROM COMPOUND INTEREST LAW (7) 1879-1913 AND FROM TWO STRAIGHT LINES, 1879-1896 AND 1897-1913 (8) AS SECULAR TRENDS.

(Equations of Secular Trends in Figures 11 and 12.)

Year.	1.(a)	2.	3.	4.	5.(b)	6.(c)	7.	8.
1879	118.2	+ 5.3	−18.5	−13.5	27	+30	− 7	− 5
1880	130.8	+17.8	− 1.8	+ 1.2	38	+33	+ 2	+ 2
1881	129.3	+16.1	+ 0.4	+ 1.8	41	+28	+ 2	+ 2
1882	132.7	+19.4	+ 7.4	+ 7.3	46	+25	+ 5	+ 3
1883	129.7	+16.2	+ 7.7	+ 6.4	46	+18	+ 2	0
1884	121.6	+ 8.0	+ 2.6	+ 0.4	41	+ 5	− 6	− 9
1885	113.8	.0	− 2.5	− 5.3	40	− 4	−10	−14
1886	112.4	− 1.5	− 1.3	− 4.6	57	+ 5	+ 3	0
1887	113.3	− .8	+ 1.9	− 1.6	64	+ 4	+ 7	+ 3
1888	115.2	+ 1.0	+ 5.8	+ 2.4	65	− 3	+ 4	+ 1
1889	115.2	+ .8	+ 7.6	+ 4.5	76	0	+11	+ 8
1890	112.9	− 1.6	+ 6.8	+ 4.3	92	+ 9	+22	+20
1891	111.7	− 3.0	+ 6.8	+ 5.2	83	− 8	+ 8	+ 8
1892	106.1	− 8.7	+ 2.2	+ 1.7	92	− 7	+12	+13
1893	105.6	− 9.4	+ 2.4	+ 3.3	71	−36	−14	−11
1894	96.1	−19.0	− 6.6	− 4.1	67	−48	−24	−19
1895	93.6	−21.7	− 8.9	− 4.5	94	−29	− 4	+ 4
1896	90.4	−25.0	−12.1	− 5.6	86	−44	−18	− 7
1897	89.7	−25.9	−13.0	− 6.5	97	−41	−14	−10
1898	93.4	−22.3	− 9.9	− 5.5	118	−28	− 1	− 1
1899	101.7	−14.1	− 2.4	+ 0.2	136	−18	+ 9	+ 5
1900	110.5	− 5.5	+ 5.3	+ 6.4	138	−24	+ 2	− 5
1901	108.5	− 7.6	+ 2.0	+ 1.8	159	−11	+14	+ 4
1902	112.9	− 3.4	+ 4.8	+ 3.6	178	0	+23	+11
1903	113.6	− 2.8	+ 3.7	+ 1.7	180	− 5	+14	+ 1
1904	113.0	− 3.6	+ 1.1	− 1.5	165	−28	−12	−26
1905	115.9	− .8	+ 1.6	− 1.2	230	+29	+41	+27
1906	122.5	+ 5.6	+ 5.6	+ 2.8	253	+44	+51	+38
1907	129.5	+12.5	+ 9.7	+ 7.2	258	+41	+42	+32
1908	122.8	+ 5.6	+ 0.1	− 2.1	159	−66	−72	−79
1909	126.5	+ 9.2	+ 0.4	− 1.0	258	+25	+11	+ 8
1910	131.6	+14.1	+ 1.8	− 1.5	273	+33	+ 9	+11
1911	129.2	+11.6	− 4.5	− 3.5	236	−12	−46	−38
1912	133.6	+15.8	− 4.1	− 1.7	297	+41	− 4	+11
1913	135.2	+17.3	− 6.9	− 2.6	310	+46	−12	+12

(a) The Aldrich and Bureau of Labor Statistics indices are reduced to a continuous series with the base 1890-1899.
(b) From Statistical Abstract of the United States, 1914, p. 664. The units here are 100,000 long tons.
(c) Equation of line of secular trend, $y = 7.852t - 3.00$, origin at 1879. This is the only trend having any negative ordinate for the period studied.

The conclusion just stated is of especial interest because it is in conflict with that of Professor H. L. Moore.* In his *Economic Cycles*, he found the following coefficients between the cycles of crop yield and pig-iron production, *using three year averages in all cases:* $r_0 = .625$; $r_{+1} = .719$; $r_{+2} = .718$; $r_{+3} = .697$; $r_{+4} = .572$ (see Table XIV). Correlating the cycles of crop yield with cycles of general prices,† he obtains the coeffi-

* Moore, H. L. *Economic Cycles*, p. 110.
† *Ibid.*, p. 122.

Digitized by Google

cients $r_{+3}=.786$, $r_{+4}=.800$, and $r_{+5}=.710$. He concludes from these coefficients, first, "that the cycles in the yield per acre of the crops are intimately related to the cycles in the activity of industry, and that it takes between one and two years for good and bad crops to produce the maximum effect upon the activity of the pig-iron industry" and, second, that "the cycles in the yield per acre of the crops are . . . . . intimately connected with the cycles of general prices, and the lag in the cycles of general prices in approximately four years."* It seems to me that this conclusion is not warranted because of the poor fit of a linear secular trend to pig-iron production. The ordinate of the secular trend is *negative* for the years 1871, 1872, and 1873. The deviations from the secular trend are all positive for the periods 1871–1877 and 1902–1910 and all negative for the period 1878–1901. The deviations of crop yield are, with few exceptions, positive from 1871 to 1879 and 1903 and 1910 and negative from 1880 to 1902.† It appears probable, then, that the correlation coefficients upon which he relies contain a large spurious element. At any rate the *differences* between the coefficients, amounting to less than .02 in most cases, on which his judgment is based, cannot be considered significant.

Waiving the question of Moore's use of three-year progressive averages to form the series for which correlation coefficients are computed, which usage throws serious doubt upon the reliability of his conclusions, I will test the correlation between the series of three-year averages by computing the coefficients between first difference of those items. The coefficients are given in Table XIV. The coefficients between (1) crop-yield and pig-iron production and between (2) crop-yield and general prices are not significant. The former group coefficients (1) shows a curious alternation in value which, examination of the basic series demonstrates, is due to a few predominating items in pig-iron production after 1905. The latter group of coefficients (2) shows a maximum at four-years lag of prices but the coefficient ($r'_{+4}=+.39$) is not much larger than the maximum coefficient ($r'_{-2}=+.32$) found for first differences of the two random series previously analyzed (see

* Moore, H. L. *Economic Cycles*, pp. 110, 122.
† *Ibid.*, p. 131.

Digitized by Google

Table V). The coefficients for pig-iron production and general prices (3) reach a maximum ($r'_{+1}=+.59$) at one year lag of prices. This coefficient is probably significant. It gives, however, a result at variance with Moore's conclusions but consistent with the conclusions obtained when the period 1879–1913 was divided into two sub-periods (see Table XIII). Moore's use of three-year averages probably results in much higher coefficients than would result from annual figures. Even so, the coefficients between first differences are small for crops and the indices of the industrial and business activity. Moore's object was to show that the cycles of business reflect the cycles in crops, not merely, having assumed that cycles exist, to find the lag. For this object a good "fit" of secular trend to data is imperative. The method of first differences, then, is valuable because it reveals spurious correlation between deviations from secular trends when the fit is not good.

TABLE XIV.

(A) COEFFICIENTS OF CORRELATION BETWEEN DEVIATIONS OF YIELD PER ACRE OF NINE CROPS FROM THE LINEAR SECULAR TREND AND SIMILAR DEVIATIONS OF PIG-IRON PRODUCTION AND OF GENERAL PRICES, 1870–1911; ITEMS OF THE VARIOUS SERIES ARE THREE-YEAR PROGRESSIVE AVERAGES.

(B) COEFFICIENTS OF CORRELATION BETWEEN FIRST DIFFERENCES OF THE RESPECTIVE DEVIATIONS.

The subscript i in r_i indicates the lag in prices and pig-iron production compared with crops, or of prices as compared with pig-iron.

(A) Moore's Coefficients. (a)

Crop Yield Correlated with:	r_{0}.	r_{+1}.	r_{+2}.	r_{+3}.	r_{+4}.	r_{+5}.
Pig-Iron Production	.625	.719	.718	.697	.572	....
General Prices	....	....	....	.786	.800	.710

(B) Coefficients between First Differences.

Crop Yield Correlated with:	r'_{0}.	r'_{+1}.	r'_{+2}.	r'_{+3}.	r'_{+4}.	r'_{+5}.
(1) Pig-Iron Production	−.02	+.26	+.06	+.25	−.12	....
(2) General Prices	−.11	−.08	+.21	+.32	+.39	+.27
Pig-Iron Production Correlated with:	r'_{-1}.	r'_{0}.	r'_{+1}.	r'_{+2}.	r'_{+3}.	r'_{+4}.
(3) General Prices	+.25	+.51	+.59	+.44	+.25	....

(a) Moore, H. L. *Economic Cycles*, pp. 110–122. Series on pp. 131, 134.

Digitized by Google

Résumé.

The variate difference correlation method has been invented to eliminate spurious correlation due to position of items in time or space.

The method involves the assumption that the taking of multiple differences leads to series of random variates. In practice for short series this assumption is not fulfilled.

Coefficients for higher differences of short series tend to alternate in sign and to conceal rather than to reveal the nature of the correlation between the series being tested.

Stability of coefficients for higher differences appears to have little significance for short series, and perhaps for long series as well. The assumption that the series correlated are made up of variates "randomly distributed in time," if fulfilled, will lead to stable coefficients for successive differences. However, though this condition is *sufficient* for stability it is not *necessary*.

In testing economic series for correspondence of their cyclical fluctuations, especially in determining the relative position of the cycles upon the assumption that there are cycles, the correlation coefficients between deviations from a linear secular trend together with coefficients for first differences constitute a reliable basis for judgment.

When the question is one of the *existence* or *non-existence* of similar cycles in two time series great care must be used in the choice of the function used to represent the secular trend and in the nature of the fit of the curve or line to the data. The method of first differences is an extremely valuable aid in investigating such a question.

Coefficients of correlation between second differences may give information concerning minor oscillations as distinct from secular trend and major cycles. Even for this purpose the use of higher than second differences appears to be unreliable, especially so for short series. The coefficients of correlation between second differences are identical with those between deviations from three-year progressive averages.

The method of measuring correlation between cycles of time series, that is both easy of application and reliable, is the method of *first* differences. In general, however, this method

Digitized by Google

should be supplemented by curve fitting. To secure a picture of the cycles, it is, of course, necessary to take deviations from a closely fitted curve.

Finally, curve fitting to eliminate the secular trend of a time series should always be adapted to the problem in hand and interpretation of coefficients of correlation between time series should be made with continual reference to the fundamental data. Important light may be secured by dividing statistical series into more homogeneous sub-series and analyzing the latter. The nature of the data is as important as the method to be applied. Rules-of-thumb concerning method or data are apt to lead to pitfalls.

Digitized by Google

SOCIAL-ECONOMIC GROUPS OF THE UNITED STATES.

Gainful Workers of United States, Classified by Social-Economic Groups or Strata.

By Alba M. Edwards, Ph.D., *Washington, D. C.*

In reporting the occupations of the gainful workers of the United States, it has been the custom of the Bureau of the Census to group the occupations into a few general divisions, each general division, as agriculture, manufactures, etc., constituting a large section of the broad field in which gainful labor is occupied. No attempt has been made by the Bureau of the Census to group the gainful workers according to social-economic groups or strata. Yet, there is a real need for such an additional grouping, for while much of our discussion and much of our labor legislation deals with the workers in a certain section or in certain sections of the industrial field, as persons engaged in agriculture, persons engaged in manufacturing and mechanical industries, etc.; another large part of our discussion—if not as yet of our legislation—deals with large social-economic groups, as proprietors, skilled workers, laborers, servants, professional persons, etc., with but minor regard to the section of the broad industrial field in which the workers in each respective group are occupied. Those discussing or desiring to discuss such social-economic groups have been hampered by the lack of any such grouping of the workers reported in the Thirteenth Census report on occupations. In the following pages such a grouping is presented.

There are those who desire a grouping of occupations according to skill. In many respects such a grouping, if it could be carried out, would be an admirable and a useful one; but a complete grouping of occupations according to skill is impossible, since many occupations do not lend themselves to such a grouping. For example, proprietors usually are distinguished from the other workers in the same industry or business, not

Digitized by Google

by a difference in skill but rather by a difference in the possession of property, credit, and business and executive ability. Where, in a scale of skill, would we place the policeman? Possibly we might, by a stretch of the imagination, classify according to skill the surgeon, but not the physician; the acrobat, but not the actor; the sculptor, but not the artist; the organist, but not the opera singer. Indeed, none of the proprietary, official, managerial, clerical, or strictly professional pursuits lends itself readily to a classification by skill; and it is doubtful whether any of them may be properly so classified, since in none of them is skill the chief characteristic. In fact, in a grouping such as here presented, we can properly classify according to skill only those occupations in which the expenditure of muscular force is one of the chief characteristics. It is impossible, of course, to draw a hard and fast line between those occupations which are characterized chiefly by the exercise of muscular force or manual dexterity, and those which are characterized chiefly by the exercise of mental force or ingenuity. In other words, it is impossible to draw a hard and fast line between the hand workers and the head workers. But such a line may be drawn sufficiently accurately for our purpose.

The grouping of the gainful workers here presented is not according to skill, except in the case of the manual workers, whose occupations lend themselves more or less readily to a classification by skill. The aim has been merely to divide the gainful workers of the United States into a few large, social-economic groups.

The grouping given below is the result of a rearrangement of the occupations and occupation groups of Table I of the Thirteenth Census report on occupations. The occupations of Table VI of the same report are in much greater detail and for this reason possibly could have been grouped with a higher degree of accuracy, but it is not believed that the added accuracy would compensate for the great additional amount of labor involved. The occupations of Table I were especially preferred for the reason that, since the occupations are classified in the same manner for each state, in Table II, and for each city of 100,000 population and over, in Table III of the

Digitized by Google

occupation report, it will be easy for any one to group in the manner here presented the occupations of any state or of any city of 100,000 population and over. And, with some estimating, a similar grouping may be made, from Table IV of the occupation report, of the occupations of any city of 25,000 to 100,000 population. Furthermore, following the classification of Table I has made possible a grouping of the occupations of the Negroes, since their occupations are classified in Table 2 of the Negro bulletin* in accordance with the classification of Table I of the occupation report.

The occupations of Table I of the Thirteenth Census report on occupations have been rearranged into the following nine groups:

I. Proprietors, officials, and managers.
II. Clerks and kindred workers.
III. Skilled workers.
IV. Semiskilled workers.
V. Laborers.
VI. Servants.
VII. Public officials.
VIII. Semiofficial public employees.
IX. Professional persons.

To any one at all familiar with occupations and occupation classification it is hardly necessary to point out the impossibility of grouping the 38,000,000 and over gainful workers of the United States into nine groups and making each group perfectly clear cut and distinct. Each of the above groups doubtless includes some workers who properly belong in another group, and from each group doubtless are omitted some workers who properly belong there. However, it is not believed that such additions and omissions are large enough to affect materially the percentage reported.

In Table I, each general division of occupations, as reported in Table I of the occupation report, is shown, with its occupations rearranged according to the grouping given above. Table I, by showing the occupations included in each group, renders a special explanation of each group unnecessary. However, since there is no unanimity of opinion among statisticians and

* "Negroes in the United States," Bureau of the Census, Bulletin 129.

Digitized by Google

others as to which occupations are skilled, which semi-skilled and which unskilled, it may be well to state briefly the rules according to which occupations were assigned, respectively, to Group III, Skilled workers, to Group IV, Semiskilled workers, and to Group V, Laborers.

As stated in a preceding paragraph, the term skill, for the purposes of a grouping such as here presented, is considered properly applied only to those occupations in which the expenditure of muscular force is one of the chief characteristics. Within this field, those occupations have been considered skilled for the pursuance of which a long period of training or an apprenticeship usually is necessary, and which in their pursuance call for a degree of judgment and manual dexterity, one or both, above that required in semiskilled occupations. Those occupations have been considered semiskilled for the pursuance of which only a short period or no period of preliminary training is necessary, and which in their pursuance call for only a moderate degree of judgment or of manual dexterity. "Laborers" have been considered to include those occupations the workers in which require no special training, judgment, or manual dexterity, but supply mainly muscular strength for the performance of coarse, heavy work.

Since, in Table I of the occupation report, the semiskilled operatives and the laborers in each kind of mines, in quarries, in oil and gas wells, and in salt wells and works were reported together, a division was necessary for the purposes of this grouping. The division is based upon estimates made from the detailed figures for each occupation of each of the industries involved, as published in Table VI of the occupation report. While these estimates are but rough ones, any probable errors in them could not affect perceptibly the percentage reported in Tables III, IV, and V, for either the "Semiskilled workers" or the "Laborers."

For the convenience of any one desiring to group similarly the occupations of any state or of any city, each occupation in Table I is preceded by its line number as published for states in Table II of the occupation report, and for cities of 100,000 population and over in Table III of the occupation report.

Digitized by Google

TABLE I.

GAINFUL WORKERS OF UNITED STATES, 10 YEARS OF AGE AND OVER, CLASSIFIED FOR EACH SEX AND FOR EACH GENERAL DIVISION OF OCCUPATIONS BY SOCIAL-ECONOMIC GROUPS OR STRATA, 1910.

Occupations.	Total.	Male.	Female.
AGRICULTURE, FORESTRY, AND ANIMAL HUSBANDRY	12,659,203	10,851,702	1,807,501
I. Proprietors, officials, and managers	6,148 387	5,874,861	273,526
37. Apiarists	2,145	2,020	125
4. Dairy farmers	61,816	59,240	2,576
6. Farmers	5,865,003	5,607,297	257,706
17. Foresters	4,332	4,332	.....
18. Gardeners, florists, fruit growers, and nurserymen	139,255	131,421	7,834
33. Owners and managers of log and timber camps	7,931	7,927	4
40. Poultry raisers and poultry yard laborers(a)	15,384	11,777	3,607
35. Stock raisers	52,521	50,847	1,674
V. Laborers	6,510,816	4,976,841	1,533,975
38. Corn shellers, hay bailers, grain threshers, etc.	5,617	5,617	.....
5. Dairy farm laborers	35,014	32,237	2,777
39. Ditchers	15,198	15,198	.....
11. Farm, dairy farm, garden, orchard, etc., foremen	47,591	39,826	7,765
7. Farm laborers	5,975,057	4,460,634	1,514,423
16. Fishermen and oystermen	68,275	67,799	476
23. Garden, greenhouse, orchard, and nursery laborers	133,927	126,453	7,474
28. Lumbermen, raftsmen, and woodchoppers	161,268	161,191	77
34. Stock herders, drovers and feeders	62,975	62,090	885
41. Other and not specified pursuits	5,894	5,796	98
EXTRACTION OF MINERALS	964,824	963,730	1,094
I. Proprietors, officials and managers	25,234	25,127	107
47. Managers	9,798	9,786	12
48. Officials	1,149	1,140	9
49. Operators	14,287	14,201	86
IV. Semiskilled workers	450,047	449,658	389
43. Foremen, overseers, and inspectors	23,338	23,328	10
50. Coal mine operatives (48 per cent. of)	294,684	294,490	194
51. Copper mine operatives (54 per cent. of)	21,206	21,196	10
52. Gold and silver mine operatives (54 per cent. of)	29,935	29,914	21
53. Iron mine operatives (51 per cent. of)	25,298	25,278	20
55. Lead and zinc mine operatives (56 per cent. of)	10,912	10,904	8
56. All other mine operatives (51 per cent. of)	14,161	14,126	35
57. Quarry operatives (22 per cent. of)	17,803	17,792	11
59. Oil and gas well operatives (46 per cent. of)	11,855	11,848	7
60. Salt well and works operatives (20 per cent. of)	855	782	73
V. Laborers	489,543	488,945	598
50. Coal mine operatives (52 per cent. of)	319,240	319,029	211
51. Copper mine operatives (46 per cent. of)	18,064	18,055	9
52. Gold and silver mine operatives (46 per cent. of)	25,501	25,483	18
53. Iron mine operatives (49 per cent. of)	24,305	24,286	19
55. Lead and zinc mine operatives (44 per cent. of)	8,574	8,567	7
56. All other mine operatives (49 per cent. of)	13,605	13,572	33
57. Quarry operatives (78 per cent. of)	63,037	63,003	34
59. Oil and gas well operatives (54 per cent. of)	13,707	13,700	7
60. Salt well and works operatives (80 per cent. of)	3,510	3,250	260

(a) Includes 3,233 poultry yard laborers.

Digitized by Google

TABLE I.—(*Continued*).

Occupations.	Total.	Male.	Female.
MANUFACTURING AND MECHANICAL INDUSTRIES	10,658,881	8,837,901	1,820,980
I. Proprietors, officials and managers	535,223	528,213	7,010
72. Builders and building contractors	174,422	173,573	849
166. Managers and superintendents manufacturing	104,210	102,748	1,462
168. Manufacturers	235,107	230,809	4,298
169. Officials	21,484	21,083	401
III. Skilled workers	3,821,327	3,750,936	70,391
66. Bakers	89,531	84,752	4,779
67. Blacksmiths, forgemen, and hammermen	240,519	240,488	31
70. Boilermakers	44,761	44,761	.....
71. Brick and stone masons	169,402	169,387	15
73. Butchers and dressers (slaughterhouse)	16,351	16,349	2
74. Cabinetmakers	41,892	41,884	8
75. Carpenters	817,120	817,082	38
76. Compositors, linotypers, and typesetters	127,589	113,538	14,051
77. Coopers	25,299	25,292	7
80. Electricians and electrical engineers	135,519	135,427	92
81. Electrotypers, stereotypers, and lithographers	12,506	11,929	577
84. Engineers (mechanical)	14,514	14,514	.....
85. Engineers (stationary)	231,041	231,031	10
86. Engravers	13,967	13,429	538
93. Glass blowers	15,564	15,474	90
95. Heaters (metal)	10,120	10,111	9
99. Jewelers, watchmakers, goldsmiths and silversmiths	32,574	30,037	2,537
96. Ladlers and pourers (metal)	679	679	.....
162. Loomfixers	13,254	13,254	.....
163. Machinists, millwrights, and toolmakers	488,049	487,956	93
170. Mechanics (n. o. s.)(a)	34,787	34,745	42
174. Millers (grain, flour, feed, etc.)	23,152	23,093	59
176. Moulders, founders, and casters (metal)	120,900	120,783	117
181. Painters, glaziers, varnishes, enamelers, etc.	337,355	334,814	2,541
185. Paper hangers	25,577	24,780	797
186. Pattern and model makers	23,559	23,006	553
187. Plasterers	47,682	47,676	6
188. Plumbers, and gas and steamfitters	148,304	148,304	.....
97. Puddlers	5,717	5,717	.....
189. Pressmen (printing)	20,084	19,892	192
190. Rollers and roll hands (metal)	18,407	18,384	23
191. Roofers and slaters	14,078	14,078	.....
192. Sawyers	43,276	43,257	19
288. Shoemakers and cobblers (not in factory)	69,570	68,788	782
289. Skilled occupations (n. o. s.)(a)	16,808	16,560	248
294. Stonecutters	35,731	35,726	5
295. Structural iron workers (building)	11,427	11,427	.....
296. Tailors and tailoresses	204,608	163,795	40,813
297. Tinsmiths and coppersmiths	59,833	59,809	24
300. Upholsterers	20,221	18,928	1,293
IV. Semiskilled workers	3,681,642	2,026,438	1,655,204
62. Apprentices	118,964	103,369	15,595
78. Dressmakers and seamstresses (not in factory)	449,342	1,582	447,760
79. Dyers	14,050	13,396	654
87. Filers, grinders, buffers and polishers (metal)	49,525	46,679	2,846
92. Foremen and overseers (manufacturing)	175,098	155,358	19,740
175. Milliners and millinery dealers	127,906	5,459	122,447
180. Oilers of machinery	14,013	13,990	23
193–286. Semiskilled operatives (n. o. s.)(a)	2,441,535	1,626,602	814,933
287. Sewers and sewing machine operatives (factory)	291,209	60,008	231,206

(a) Not otherwise specified.

Digitized by Google

TABLE I.—(*Continued*).

Occupations.	Total.	Male.	Female.
MANUFACTURING AND MECHANICAL INDUSTRIES—*Continued.*			
V. Laborers	2,620,689	2,532,314	88,375
91. Firemen (except locomotive and fire department)	111,248	111,248	.....
94. Furnacemen and smeltermen	19,735	19,719	16
103–161. Laborers (n. o. s.)	2,489,706	2,401,347	88,359
TRANSPORTATION	2,637,671	2,531,075	106,596
I. Proprietors, officials, and managers	182,595	180,729	1,866
303. Captains, masters, mates, and pilots	24,242	24,242	.....
319. Conductors (steam railroad)	65,604	65,604	.....
310. Garage keepers and managers	5,279	5,256	23
312. Livery stable keepers and managers	34,795	34,612	183
328. Officials and superintendents	22,238	22,236	2
313. Proprietors and managers of transfer companies	15,598	15,368	230
358. Proprietors, officials, and managers (n. o. s.)(a)	14,839	13,411	1,428
II. Clerks and kindred workers	326,743	227,886	98,857
336. Agents (express companies)	5,875	5,804	71
314. Baggage men and freight agents	17,033	17,028	5
337. Express messengers and railway mail clerks	22,021	22,018	3
340. Mail carriers	80,678	79,667	1,011
342. Telegraph messengers	9,152	9,074	78
343. Telegraph operators	69,953	61,734	8,219
344. Telephone operators	97,893	9,631	88,262
335. Ticket and station agents	24,138	22,930	1,208
III. Skilled workers	200,271	200,135	136
351. Inspectors "steam railroad"	27,661	27,525	136
325. Locomotive engineers	96,229	96,229	.....
326. Locomotive firemen	76,381	76,381	.....
IV. Semiskilled workers	572,586	570,689	1,897
302. Boatmen, canal men, and lock keepers	5,304	5,289	15
317. Boiler washers and engine hostlers	10,409	10,409	.....
318. Brakemen	92,572	92,572	.....
306. Carriage and hack drivers	35,376	35,339	37
307. Chauffeurs	45,785	45,752	33
320. Conductors (street railroad)	56,932	56,932	.....
321. Foremen and overseers (railroads)	69,933	69,693	240
345. Foremen and overseers (n. o. s.)(a)	14,738	14,333	405
309. Foremen of livery and transfer companies	6,606	6,606	.....
353. Inspectors, "other transportation"	3,308	3,172	136
352. Inspectors, "street railroad"	2,268	2,265	3
327. Motormen	59,005	59,005	.....
305. Sailors and deck hands	46,510	46,498	12
331. Switchmen, flagmen, and yardmen	85,147	85,095	52
361. Other occupations (semiskilled)	38,693	37,729	964
V. Laborers	1,355,476	1,351,636	3,840
308. Draymen, teamsters and expressmen	408,469	408,396	73
311. Hostlers and stable hands	63,388	63,382	6
322. Laborers (railroad)	570,975	567,522	3,453
354. Laborers (n. o. s.)(a)	221,437	221,176	261
304. Longshoremen and stevedores	62,857	62,813	44
341. Telegraph and telephone linemen	28,350	28,347	3

(a) Not otherwise specified.

Digitized by Google

TABLE I.—(*Continued*).

Occupations.	Total.	Male.	Female.
TRADE	3,614,670	3,146,582	468,088
I. Proprietors, officials, and managers	1,404,478	1,331,868	72,610
366. Bankers, brokers, and money lenders	105,804	103,170	2,634
385. Officials of insurance companies	9,501	9,376	125
394. Proprietors, officials, and managers (n. o. s.)(a)	22,362	21,352	1,010
400. Retail dealers	1,195,029	1,127,926	67,103
406. Undertakers	20,734	19,921	813
407. Wholesale dealers, importers, and exporters	51,048	50,123	925
II. Clerks and kindred workers	1,717,650	1,335,472	382,178
373. Clerks in stores	387,183	275,589	111,594
374. Commercial travelers	163,620	161,027	2,593
380. Floorwalkers and foremen in stores	17,946	14,900	3,046
382. Inspectors, gaugers, and samplers	13,446	11,685	1,761
384. Insurance agents	88,463	85,926	2,537
399. Real estate agents and officials	125,862	122,935	2,927
401. Salesmen and saleswomen	921,130	663,410	257,720
IV. Semiskilled workers	309,086	300,623	8,463
375. Decorators, drapers, and window dressers	5,341	4,902	439
376. Deliverymen	229,619	229,469	150
381. Foremen, warehouses, stock yards, etc.	2,778	2,749	29
393. Newsboys	29,708	29,435	273
406. Other pursuits (semiskilled)	41,640	34,068	7,572
V. Laborers	183,456	178,619	4,837
386. Laborers in coal and lumber yards, warehouses, etc.	81,123	80,450	673
392. Laborers, porters, and helpers in stores	102,333	98,169	4,164
PUBLIC SERVICE (not elsewhere classified)	459,291	445,733	13,558
VIII. Public officials	128,779	116,276	12,503
418. Marshals, sheriffs, and detectives	23,599	23,219	380
423. Officials and inspectors (city and county)	52,254	49,668	2,586
426. Officials and inspectors (state and United States)	52,926	43,389	9,537
VIII. Semiofficial public employees (not elsewhere classified)	263,278	262,952	326
413. Firemen, fire department	35,606	35,606	.....
414. Guards, watchmen, and doorkeepers	78,271	78,168	103
432. Life savers	2,158	2,158	.....
433. Lighthouse keepers	1,593	1,552	41
429. Policemen	61,980	61,980	.....
430. Soldiers, sailors, and marines	77,153	77,153	.....
434. Other occupations	6,517	6,335	182
V. Laborers	67,234	66,505	729
415. Laborers (public service)	67,234	66,505	729
PROFESSIONAL SERVICE	1,663,569	929,684	733,885
IX. Professional persons	1,644,968	919,369	725,599
(All of "Professional service," except 473—"Attendants, and helpers professional service.")			
IV. Semiskilled workers	18,601	10,315	8,286
473. Attendants and helpers (professional service)	18,601	10,315	8,286

(a) Not otherwise specified.

Digitized by Google

TABLE I.—(*Concluded*).

Occupations.	Total.	Male.	Female.
DOMESTIC AND PERSONAL SERVICE	3,772,174	1,241,328	2,530,846
I. Proprietors, officials, and managers	393,807	223,361	170,446
477. Billiard room, dance hall, skating rink, etc., keepers	16,761	15,943	818
480. Boarding and lodging house keepers	165,452	23,052	142,400
484. Hotel keepers and managers	64,504	50,269	14,235
490. Laundry owners, officials and managers	18,043	17,057	986
495. Restaurant, café, and lunch room keepers	60,832	50,316	10,516
496. Saloon keepers	68,215	66,724	1,491
IV. Semiskilled workers	659,140	265,965	393,175
475. Barbers, hairdressers, and manicurists	195,275	172,977	22,298
485. Housekeepers and stewards	189,273	15,940	173,333
489. Laundry operatives	111,879	35,899	75,980
491. Midwives and nurses (not trained)	133,043	15,926	117,117
504. Other pursuits	29,670	25,223	4,447
VI. Servants	2,719,227	752,002	1,967,225
476. Bartenders	101,234	100,984	250
481. Bootblacks	14,020	14,000	20
482. Charwomen and cleaners	34,034	7,195	26,839
483. Elevator tenders	25,035	25,010	25
486. Janitors and sextons	113,081	91,629	21,452
487. Laborers (domestic and professional service)	53,480	50,265	3,215
488. Launderers and laundresses (not in laundries)	533,697	13,693	520,004
494. Porters (except in stores)	84,128	84,055	73
497. Servants	1,572,225	262,676	1,309,549
503. Waiters	188,293	102,495	85,798
CLERICAL OCCUPATIONS	1,737,053	1,143,829	593,224
II. Clerks and kindred workers	1,737,053	1,143,829	593,224
511-522. All classed under "Clerical occupations."			

Table II, in which are brought together the results of Table I, shows the number of workers in each of nine social-economic groups, and, for each group, a distribution of the workers by general division of occupations.

Table III shows for the total workers and for the workers of each sex the number and the proportion of persons in each of nine social-economic groups. According to this table, 22.8 per cent. of the gainful workers of the United States, in 1910, were proprietors, officials, or managers. Almost one worker in every ten (9.9 per cent.) was a clerk or kindred worker, and slightly more than one in every ten (10.5 per cent.) was a skilled worker. Semiskilled workers formed 14.9 per cent., laborers 29.4 per cent., and servants 7.1 per cent. of the workers. Public officials, semiofficial public employees, and professional persons together constituted only 5.3 per cent. of the workers.

Digitized by Google

TABLE II.

GAINFUL WORKERS OF THE UNITED STATES, 10 YEARS OF AGE AND OVER, CLASSIFIED BY SOCIAL-ECONOMIC GROUPS OR STRATA, 1910.

Groups.	Total.	Male.	Female.
ALL GAINFUL WORKERS	38,167,336	30,091,564	8,075,772
I. Proprietors, officials, and managers	8,689,724	8,164,159	525,565
Agriculture, forestry and animal husbandry	6,148,387	5,874,861	273,526
Extraction of minerals	25,234	25,127	107
Manufacturing and mechanical industries	535,223	528,213	7,010
Transportation	182,595	180,729	1,866
Trade	1,404,478	1,331,868	72,610
Domestic and personal service	393,807	223,361	170,446
II. Clerks and kindred workers	3,781,446	2,707,187	1,074,259
Transportation	326,743	227,886	98,857
Trade	1,717,650	1,335,472	382,178
Clerical occupations	1,737,053	1,143,829	593,224
III. Skilled workers	4,021,598	3,951,071	70,527
Manufacturing and mechanical industries	3,821,327	3,750,936	70,391
Transportation	200,271	200,135	136
IV. Semiskilled workers	5,691,102	3,623,688	2,067,414
Extraction of minerals	450,047	449,658	389
Manufacturing and mechanical industries	3,681,642	2,026,438	1,655,204
Transportation	572,586	570,689	1,897
Trade	309,086	300,623	8,463
Professional service	18,601	10,315	8,286
Domestic and personal service	659,140	265,965	393,175
V. Laborers	11,227,214	9,594,860	1,632,354
Agriculture, forestry, and animal husbandry	6,510,816	4,976,841	1,533,975
Extraction of minerals	489,543	488,945	598
Manufacturing and mechanical industries	2,620,689	2,532,314	88,375
Transportation	1,355,476	1,351,636	3,840
Trade	183,456	178,619	4,837
Public service	67,234	66,505	729
VI. Servants	2,719,227	752,002	1,967,225
Domestic and personal service	2,719,227	752,002	1,967,225
VII. Public officials	128,779	116,276	12,503
Public service	128,779	116,276	12,503
VIII. Semiofficial public employees	263,278	262,952	326
Public service	263,278	262,952	326
IX. Professional persons	1,644,968	919,369	725,599
Professional service	1,644,968	919,369	725,599

In this grouping, the distinction between the sexes is quite marked. While 27.1 per cent. of the male workers are in the proprietary, official, and managerial group, only 6.5 per cent. of the female workers are in this group; and while 13.1 per cent. of the males are skilled workers, only 0.9 per cent. of the females are skilled workers. Only 2.5 per cent. of the male workers are servants, as compared with 24.4 per cent. of the females; and only 3.1 per cent. of the male workers are profes-

Digitized by Google

sional persons, as compared with 9 per cent. of the females. Semiskilled workers, laborers, and servants together constitute over seven out of ten (70.2 per cent.) of the female workers.

In Table I of the Thirteenth Census occupation report, from the occupations of which the groups here presented were formed, certain specific occupations which, technically, are skilled occupations were classified as semiskilled because the enumerators returned so many children, young persons, and women as pursuing these occupations as to render the occupations semiskilled, even though each of them did contain some skilled workers. For this reason, it is believed that the group of skilled workers as here presented is somewhat too small.

TABLE III.

NUMBER AND PROPORTION OF THE TOTAL WORKERS AND OF THE WORKERS OF EACH SEX ENGAGED IN EACH OF NINE SOCIAL-ECONOMIC GROUPS, UNITED STATES, 1910.

Group.	Total.		Male.		Female.	
	Number.	Per Cent.	Number.	Per Cent.	Number.	Per Cent.
ALL GROUPS	38,167,336	100.0	30,091,564	100.0	8,075,772	100.0
I. Proprietors, officials and managers	8,689,724	22.8	8,164,159	27.1	525,565	6.5
II. Clerks and kindred workers	3,781,446	9.9	2,707,187	9.0	1,074,259	13.3
III. Skilled workers	4,021,598	10.5	3,951,071	13.1	70,527	0.9
IV. Semiskilled workers	5,691,102	14.9	3,623,688	12.0	2,067,414	25.6
V. Laborers	11,227,214	29.4	9,594,860	31.9	1,632,354	20.2
VI. Servants	2,719,227	7.1	752,002	2.5	1,967,225	24.4
VII. Public officials	128,779	0.3	116,276	0.4	12,503	0.2
VIII. Semiofficial public employees	263,278	0.7	262,952	0.9	326	(a)
IX. Professional persons	1,644,968	4.3	919,369	3.1	725,599	9.0

(a) Less than one tenth of 1 per cent.

Table IV shows for the total Negro workers of the United States, and for the Negro workers of each sex, the number and proportion in each of nine social-economic groups in 1910. For purposes of comparison, the data for the total workers, given in Table III, and the data for the Negro workers, given in Table IV, are reproduced in Table V, with additional data for "White and all other" workers. Table V thus shows, for both sexes and for each sex separately, the number and proportion of the total workers, of the Negro workers, and of the white and all other workers engaged in each of nine social-economic groups in 1910. Since only 200,475 of the workers

Digitized by Google

in the "White and all other" group are nonwhite, this group, for comparative purposes, may be considered practically white. Therefore, in the discussion of Table V, the workers in this group will be referred to as white.

TABLE IV.

NUMBER AND PROPORTION OF THE TOTAL NEGRO WORKERS AND OF THE NEGRO WORKERS OF EACH SEX ENGAGED IN EACH OF NINE SOCIAL-ECONOMIC GROUPS, UNITED STATES, 1910.

Group.	Total.		Male.		Female.	
	Number.	Per Cent.	Number.	Per Cent.	Number.	Per Cent.
ALL GROUPS	5,192,535	100.0	3,178,554	100.0	2,013,981	100.0
I. Proprietors, officials, and managers	933,538	18.0	837,872	26.4	95,666	4.8
II. Clerks and kindred workers	36,496	0.7	30,386	1.0	6,110	0.3
III. Skilled workers	112,708	2.2	111,852	3.5	856	(a)
IV. Semiskilled workers	274,585	5.3	172,965	5.4	101,620	5.0
V. Laborers	2,732,161	52.6	1,746,227	54.9	985,934	49.0
VI. Servants	1,027,800	19.8	233,181	7.3	794,619	39.5
VII. Public officials	923	(a)	831	(a)	92	(a)
VIII. Semiofficial public employees	8,454	0.2	8,435	0.1	19	(a)
IX. Professional persons	65,870	1.3	36,805	1.2	29,065	1.4

(a) Less than one tenth of 1 per cent.

Coming at once, in Table V, to the percentages for the male workers, we note that 26.4 per cent., of the Negroes, as compared with 27.2 per cent. of the whites, were proprietors, officials, and managers. The fact that almost as large a proportion of the Negroes as of the whites were in this group is explained by the further fact that 25.3 per cent. of all Negro, as compared with 18.8 per cent. of all white male workers were agricultural proprietors. Agricultural proprietors constituted 96 per cent. of the Negro and but 69.2 per cent. of the white male workers in Group I. In 1910, but a small proportion of the Negro male workers were engaged in clerical or kindred pursuits—only 1 per cent. of them being in this group, as compared with 9.9 per cent. of the whites. Likewise, the proportion of the Negro males who were skilled workers, 3.5 per cent., was quite small as compared with the proportion for the white males, 14.3 per cent.; and the Negro semiskilled male workers constituted only 5.4 per cent. of the total, as compared with 12.8 per cent. for the white semiskilled male workers. Of the

Digitized by Google

Negro male workers, considerably more than one half were laborers—54.9 per cent. as compared with 29.2 per cent. for the white male workers. Of the Negro male laborers, 59.5 per cent. were agricultural laborers, etc., as compared with 37.8 per cent. of the white male laborers. Servants constituted 7.3 per cent. of all Negro male workers, and but 1.9 per cent. of all white male workers. Laborers and servants combined formed 62.2 per cent. of the Negro male workers, and but one half as large a proportion (31.1 per cent.) of the white male workers. The proportion of the male workers who were public officials, semiofficial public employees, and professional persons, respectively, was very much smaller for the Negroes than for the whites.

Of the female gainful workers, 4.8 per cent. of the Negroes and 7.1 per cent. of the whites were proprietors, officials, or managers. Only 0.3 per cent. of the Negro female workers were engaged in clerical or kindred pursuits, as compared with 17.6 per cent. of the white female workers. Semiskilled workers constituted 5 per cent. of all Negro female workers and 32.4 per cent. of all white female workers. Almost one half (49 per cent.) of the Negro female workers were laborers, and almost two out of every five of them (39.5 per cent.) were servants. Laborers and servants together constituted not far from nine out of every ten (88.5 per cent.) of the Negro female workers, as compared with only three out of every ten (30 per cent.) of the white female workers. Only 1.4 per cent. of the Negro female workers were engaged in professional pursuits, as compared with 11.5 per cent. of the white female workers.

Table VI shows the number and the proportion of the gainful workers in each specified social-economic group and subgroup at each of the censuses, 1910, 1900, 1890, 1880, and 1870. The classification of occupations followed at the Thirteenth Census shows occupations in so much greater detail than they were shown at preceding censuses that it was impossible in certain cases to rearrange the occupations of preceding censuses according to the grouping given in Tables III, IV, and V for the Thirteenth Census occupations. In Table VI, it has been necessary to combine groups III, IV, and V of Tables

Digitized by Google

III, IV, and V—the skilled workers, the semiskilled workers, and the laborers.

TABLE V.

NUMBER AND PROPORTION OF THE TOTAL WORKERS OF THE UNITED STATES, OF THE NEGRO WORKERS, AND OF THE WHITE AND ALL OTHER WORKERS ENGAGED IN EACH OF NINE SOCIAL-ECONOMIC GROUPS, FOR BOTH SEXES AND FOR EACH SEX SEPARATELY, 1910.

Group and Sex.	Total.		Negro.		White and all Other.(a)	
	Number.	Per Cent.	Number.	Per Cent.	Number.	Per Cent.
BOTH SEXES.						
ALL GROUPS	38,167,336	100.0	5,192,535	100.0	32,974,801	100.0
I. Proprietors, officials, and managers	8,689,724	22.8	933,538	18.0	7,756,186	23.5
Agriculture, forestry, and animal husbandry	6,148,387	16.1	883,979	17.0	5,264,408	16.0
Other general divisions	2,541,337	6.7	49,559	1.0	2,491,778	7.6
II. Clerical and kindred workers	3,781,446	9.9	36,496	0.7	3,744,950	11.4
III. Skilled workers	4,021,598	10.5	112,708	2.2	3,908,890	11.9
IV. Semiskilled workers	5,691,102	14.9	274,585	5.3	5,416,517	16.4
V. Laborers	11,227,214	29.4	2,732,161	52.6	8,495,053	25.8
VI. Servants	2,719,227	7.1	1,027,900	19.8	1,691,427	5.1
VII. Public officials	128,779	0.3	923	(b)	127,856	0.4
VIII. Semiofficial public employees	263,278	0.7	8,454	0.2	254,824	0.8
IX. Professional persons	1,644,968	4.3	65,870	1.3	1,579,098	4.8
MALES.						
ALL GROUPS	30,091,564	100.0	3,178,554	100.0	26,913,010	100.0
I. Proprietors, officials, and managers	8,164,159	27.1	837,872	26.4	7,326,287	27.2
Agriculture, forestry, and animal husbandry	5,874,861	19.5	804,004	25.3	5,070,857	18.8
Other general divisions	2,289,298	7.6	33,868	1.1	2,255,430	8.4
II. Clerical and kindred workers	2,707,187	9.0	30,386	1.0	2,676,801	9.9
III. Skilled workers	3,951,071	13.1	111,852	3.5	3,839,219	14.3
IV. Semiskilled workers	3,623,688	12.0	172,965	5.4	3,450,723	12.8
V. Laborers	9,594,860	31.9	1,746,227	54.9	7,848,633	29.2
VI. Servants	752,002	2.5	233,181	7.3	518,821	1.9
VII. Public officials	116,276	0.4	831	(b)	115,445	0.4
VIII. Semiofficial public employees	262,952	0.9	8,435	0.1	254,517	0.9
IX. Professional persons	919,369	3.1	36,806	1.2	882,564	3.3
FEMALES.						
ALL GROUPS	8,075,772	100.0	2,013,981	100.0	6,061,791	100.0
I. Proprietors, officials, and managers	525,565	6.5	95,666	4.8	429,899	7.1
Agriculture, forestry, and animal husbandry	273,526	3.4	79,975	4.0	193,551	3.2
Other general divisions	252,039	3.1	15,691	0.8	236,348	3.9
II. Clerical and kindred workers	1,074,259	13.3	6,110	0.3	1,068,149	17.6
III. Skilled workers	70,527	0.9	856	(b)	69,671	1.1
IV. Semiskilled workers	2,067,414	25.6	101,620	5.0	1,965,794	32.4
V. Laborers	1,632,354	20.2	985,934	49.0	646,420	10.7
VI. Servants	1,967,225	24.4	794,619	39.5	1,172,606	19.3
VII. Public officials	12,503	0.2	92	(b)	12,411	0.2
VIII. Semiofficial public employees	326	(b)	19	(b)	307	(b)
IX. Professional persons	725,599	9.0	29,065	1.4	696,534	11.5

(a) Only six tenths of 1 per cent. of this group is nonwhite.
(b) Less than one tenth of 1 per cent.

Digitized by Google

TABLE VI.

NUMBER AND PROPORTION OF THE GAINFUL WORKERS OF THE UNITED STATES IN EACH SPECIFIED SOCIAL-ECONOMIC GROUP AND SUB-GROUP, 1910, 1900, 1890, 1880, 1870.(a)

Groups.	1910.		1900.		1890.		1880.		1870.	
	Number.	Per Cent.	Number.	Per Cent.	Number.	Per Cent.	Number.	Per Cent.	Number.	Per Cent.
ALL GROUPS	38,167,336	100.0	29,073,233	100.0	23,318,183	100.0	17,392,099	100.0	12,505,923	100.0
I. Proprietors, officials and managers	8,689,724	22.8	7,548,969	26.0	6,723,746	28.8	5,167,558	29.7	3,650,907	29.2
1. Agriculture	6,148,387	16.1	5,757,757	19.8	5,374,349	23.0	4,301,202	24.7	3,025,407	24.2
2. Manufacturing (including mining, quarrying, etc.)	560,457	1.5	240,651	0.8	162,386	0.7	68,936	0.4	56,844	0.5
3. Trade and transportation	1,587,073	4.2	1,295,284	4.5	1,009,401	4.3	689,377	4.0	494,655	4.0
4. Domestic and personal service	393,807	1.0	255,277	0.9	187,610	0.8	108,043	0.6	74,001	0.6
II. Clerks and kindred workers	3,781,446	9.9	2,143,353	7.4	1,387,326	5.9	655,647	3.8	366,386	2.9
III, IV, and V. Skilled and semiskilled workers and laborers	20,939,914	54.9	15,823,183	54.4	12,363,364	53.0	9,608,886	55.2	7,037,810	56.3
1. Agricultural laborers, etc.	6,510,816	17.1	4,733,197	16.3	3,866,701	16.6	3,483,497	20.0	2,966,282	23.7
2. Industrial manual workers	13,684,123	35.9	10,499,941	36.1	8,149,224	34.9	5,884,781	33.8	3,917,637	31.3
(a) Extraction of minerals	939,590	2.5	581,212	2.0	396,805	1.7	257,312	1.5	170,635	1.4
(b) Manufacturing	10,123,658	26.5	7,930,928	27.3	6,334,907	27.2	4,643,678	26.7	3,105,259	24.8
(c) Trade and transportation	2,620,875	6.9	1,987,801	6.8	1,417,512	6.1	983,791	5.7	641,743	5.1
3. Semiskilled workers and laborers in service(b)	744,975	2.0	590,045	2.0	347,439	1.5	237,608	1.4	153,891	1.2
VI. Servants	2,719,227	7.1	2,164,924	7.4	1,828,681	7.8	1,311,160	7.5	1,045,288	8.4
VII. Public officials	128,779	0.3	82,975	0.3	79,120	0.3	66,827	0.4	45,046	0.4
VIII. Semiofficial public employees	263,278	0.7	169,883	0.6	97,761	0.4	65,728	0.4	36,756	0.3
IX. Professional persons	1,644,968	4.3	1,139,946	3.9	838,185	3.6	519,298	3.0	323,730	2.6

(a) The contents of each group of Table VI are shown by Table I.

(b) Public, professional, and domestic and personal service.

Digitized by Google

For some purposes it is desirable to have—in addition to the skilled workers, the semiskilled workers, and the laborers—another group, commonly, though somewhat incorrectly, called "industrial wage earners." This group includes the manual workers engaged in the extraction of minerals, manufacturing, transportation, and trade. Since many skilled workmen, such as carpenters, painters, etc., who are working on their own account and are not receiving wages, are included in this group, it is not exactly accurate to call the group "industrial wage earners." Industrial manual workers is a better designation. The group may be formed by combining, in Table II, above, all the skilled workers, and the semiskilled workers and the laborers in the extraction of minerals, manufacturing and mechanical industries, transportation, and trade. This group is brought out in Table VI as one of the three sub groups of group "III, IV, and V."

For the purposes of the grouping for Table VI, it was necessary in three different cases to separate occupations combined at some of the earlier censuses. In making these separations, each occupation was allotted that portion of the total for the combination, which it constituted at the first census it was shown separately. "Laborers (not specified)," classified under domestic and personal service prior to the Thirteenth Census, were distributed for 1900, 1890, 1880, and 1870, respectively, by allotting to each of four groups of Table VI,—agricultural laborers, etc., industrial manual workers in manufacturing, industrial manual workers in trade and transportation, and semiskilled workers and laborers in service—that portion of the total which a careful analysis of the occupational designations classified in 1900 under "Laborers (not specified)" indicated the group contained in 1910. Also, because of the differences in classification, some estimating was necessary in the case of each group of Table VI in order to avoid having two slightly different numbers and per cents. for each group for the census year 1910, one number and one per cent. in Tables III and V and another and slightly different number and per cent. in Table VI. But, since in the case of no group did the number involved in the estimate equal as many as 1 per cent. of the total workers for the census year, it

Digitized by Google

is not probable that errors in these estimates affected materially the percentage reported for any group. And, since, for example, an error of 381,673 would be necessary in order to affect as much as 0.1 per cent. the percentage reported for any group for the year 1910, it is not probable that any percentage reported has been affected materially by the different estimates made.

According to Table VI, the proportion of the gainful workers of the United States engaged in proprietary, official and managerial positions (Group I) decreased from 29.2 per cent. in 1870 to 22.8 per cent. in 1910. This decrease was confined entirely to agricultural proprietors who constituted 24.2 per cent. of all gainful workers in 1870 and only 16.1 per cent. in 1910. The proportion which the workers in each of the other subgroups of Group I constituted of all gainful workers was considerably larger in 1910 than in 1870.

Clerical and kindred workers increased rapidly from 2.9 per cent. of all workers in 1870 to 9.9 per cent. in 1910. No other group made so great a change during this period in the proportion it constituted of the total gainful workers. The numerical increase in the workers engaged in each clerical or kindred pursuit was also quite general.

It is interesting to note that between 1870 and 1910 there was no marked change in the proportion which the large group, the skilled and semiskilled workers and the laborers, combined (Group III, IV, and V), constituted of all gainful workers. During the four decades this proportion did not vary greatly from 55 per cent. In certain of the sub groups of Group III, IV, and V, however, the proportion which the workers constituted of the total changed considerably between 1870 and 1910. "Agricultural laborers, etc." decreased during this period from 23.7 per cent. to 17.1 per cent. of the total. This decrease was in line with, though not so rapid as, the decrease in the proportion which agricultural proprietors constituted of the total workers. The increase between 1900 and 1910 in the proportion which "Agricultural laborers, etc." constituted of all workers is believed to be due to the enumeration as agricultural laborers, in 1910, of women and children such as would not have been so enumerated in

Digitized by Google

1890.* The "Industrial manual workers" (Group III, IV, and V, sub group 2) are the workers in whose interest most of our labor and much of our social legislation has been and is being enacted. This is a group consisting mainly of industrial wage earners. It is, for many reasons, a group of peculiar and special interest. It increased from 31.3 per cent. of all workers in 1870 to 35.9 per cent. in 1910. There was, also, during this period, an increase in the proportion which the workers in each sub group of the industrial manual workers constituted of all gainful workers. The industrial manual workers may be classified according to skill for the census of 1910 as follows:

INDUSTRIAL MANUAL WORKERS, 1910

Sex and Group.	Number.	Per Cent. of	
		Industrial Manual Workers.	Total Gainful Workers.
BOTH SEXES.			
Total	13,684,123	100.0	35.9
Skilled workers	4,021,598	29.4	10.5
Semiskilled workers	5,013,361	36.6	13.1
Laborers	4,649,164	34.0	12.2
MALE.			
Total	11,849,993	100.0	39.4
Skilled workers	3,951,071	33.3	13.1
Semiskilled workers	3,347,408	28.2	11.1
Laborers	4,551,514	38.4	15.1
FEMALE.			
Total	1,834,130	100.0	22.7
Skilled workers	70,527	3.8	0.9
Semiskilled workers	1,666,953	90.8	20.6
Laborers	97,650	5.3	1.2

While the number of persons in the servant group considerably more than doubled during the 40 years from 1870 to 1910, the proportion which servants of all kinds constituted of the total gainful workers declined from 8.4 per cent. in 1870 to 7.1 per cent. in 1910. This group, as shown by Group VI of Table I, above, includes all servant pursuits. Servants

* See Thirteenth Census report on occupations, pp. 26–29.

Digitized by Google

and waiters, more strictly defined, decreased, relatively, even more rapidly—from 7.5 per cent. (941,392) of all gainful workers, in 1870, to 4.9 per cent. (1,867,448) in 1910.

The proportion which public officials (Group VII) constituted of all gainful workers declined slightly between 1870 and 1910; but the semiofficial public employees* increased from 0.3 per cent. of all gainful workers in 1870 to 0.7 per cent. in 1910. Professional persons increased rapidly in relative numerical importance from 2.6 per cent. of all gainful workers in 1870 to 4.3 per cent. in 1910.

* Firemen, fire department; guards, watchmen, and doorkeepers; life savers; lighthouse keepers; policemen; soldiers, sailors, and marines; and "other occupations."

Digitized by Google

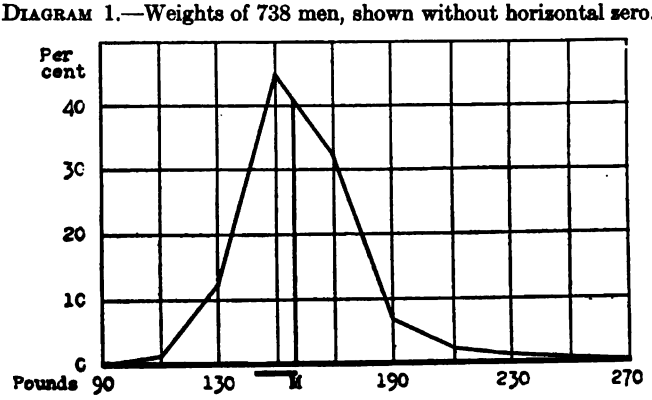
THE HORIZONTAL ZERO IN FREQUENCY DIAGRAMS.

By Earle Clark, *Russell Sage Foundation.*

It is a generally accepted rule of graphic presentation that a zero, used in a diagram as a point of reference, should be included in the diagram. This rule, while it is observed in most statistical work, is almost universally disregarded in the drafting of frequency diagrams.

Diagram 1, presented herewith, is a frequency graph of a common type, based on the weights of 738 men.* Weights are indicated on the base line, and the per cent. of cases cor-

Diagram 1.—Weights of 738 men, shown without horizontal zero.



responding to any given weight is proportionate to the vertical distance from the base line to the curve. A zero line is the most conspicuous feature of this diagram, but inspection of the figure shows that the presentation implies two zeros, and that only one of these is shown. The vertical scale, representing percentages, begins at the zero base line, but the horizontal scale, representing weights, begins at 90 pounds. It is the purpose of this paper to state reasons for including the horizontal zero, to direct attention to a type of frequency

*The data are for 738 men born in Wales, as shown in Yule's "Introduction to the Theory of Statistics," p. 95. For convenience in presentation, the extremes of the distribution have been arbitrarily shortened.

Digitized by Google

diagrams to which these reasons do not apply, and to illustrate methods of drafting.

A frequency diagram is plotted for the purpose of showing the significant facts about a series of variables. The graphic form is used rather than a frequency table or text statement because most people, even most statisticians, find it easier to perceive and appreciate these significant facts by looking at a diagram than by studying a column of figures. The essential facts about a variable series are: (1) the mean, median, or other measure of central tendency, and (2) the distribution of the values about this central tendency. These facts are interdependent. It is a simple matter to compute medians or means, but these measures do not reveal the whole truth about a distribution; they may be seriously misleading unless shown in relation to the distribution of the individual values.

On the other hand, the distribution is not in itself significant unless related to the central tendency. Stated in pounds and ounces, the average deviation* of the weights of a group of 1,000 elephants would doubtless be far greater than the average deviation of the weights of 1,000 canary birds, but this would not necessarily mean that the weights of elephants are relatively more variable than the weights of canary birds. In order to determine the true variability of a series it is necessary to relate the measure of dispersion to the measure of central tendency. This may be done by computing a coefficient of dispersion†—a ratio which expresses the dispersion as a proportion of the measure of central tendency.

It follows that, if a frequency diagram is to serve the purpose for which it is intended, it must show, with all possible clearness and effectiveness, the distribution of the individual values, the central tendency, and the relation of the distribution to the central tendency. Diagram 1 shows the distribution of the measures. Does it also show, with the emphasis required, the two other essential facts?

On Diagram 1 the median is indicated in the usual way—

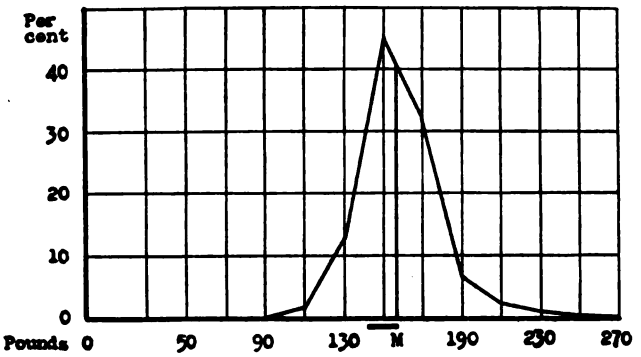
*The average deviation is a statistical expression indicating the dispersion of the values of a series of variables about their central tendency. It is obtained by adding together the differences between all the individual values and the central tendency, and dividing the result by the number of values.

†The coefficient of dispersion is the measure of dispersion (the average deviation, standard deviation, or probable error) divided by the measure of central tendency.

Digitized by Google

by a vertical line dividing into two equal parts the surface of the figure enclosed by the curve and the base line. This line is sometimes referred to as the median line, but the designation does violence to the principles of graphic presentation. In diagrams, lines or areas are, or should be, proportionate to the quantities they represent. The length of the so-called "median line" is not proportionate to the median weight of men; it is proportionate rather, as the class interval for the distribution is 20 pounds, to the approximate number of men whose weights fall within limits fixed, respectively, at 10 pounds below and at 10 pounds above the median weight. The line represents, in other words, not the median value for the series,

DIAGRAM 2.—Weights of 738 men, shown with horizontal zero.



but a number of cases. There is nowhere on the diagram a line representing by its length, or a surface representing by its area, the median weights of the men.

The median can be determined, it is true, by referring to the scale at the foot of the figure. As the point of intersection of the so-called "median line" with the base line falls at 156 pounds, as indicated by the horizontal scale, it follows that this value is the median, but the result is not obtained by the graphic method. The figures on the scale are not graphic representations any more than are the figures of a table or a text statement.

The median can, however, be shown by the graphic method by so extending the base line that the horizontal scale will include the zero. This method has been followed in pre-

Digitized by Google

paring Diagram 2. In Diagram 2 the horizontal distance from the vertical line at the left of the figure to the so-called "median line," measured on the base line or along any abscissa, represents the median weight of the men.

If the inclusion of the horizontal zero is required for a complete graphical representation of the median, it is even more essential as a means of showing the relationship of the dispersion to the median. As Diagram 1 contains no graphical representation of the central tendency, it follows that it affords no graphical representation of the relation between the central tendency and the dispersion. The dispersion of the series is indicated by the form of the curve and also by a line beneath the base line, proportionate in length to the average deviation (14.2 pounds), drawn to scale and extending to the left of the median. By including this line, the dispersion is reduced to a single graphical expression, but the diagram contains no graphical representation of the median with which either the line or the curve can be compared.

An effective graphical representation of the relationship between the central tendency and the distribution is found in Diagram 2, in which the median, represented by the distance between the horizontal zero and the vertical "median line," can be compared both with the surface of frequency, as indicated by the curve, and with the line representing the average deviation. The ratio of the length of this line to the distance from the horizontal zero to the median line is equivalent to the coefficient of dispersion.

The difficulties arising from the omission of the horizontal zero are further illustrated in Diagram 3, in which the weights of the 738 men are compared with the weights of 279 thirteen- and fourteen-year-old school boys.*

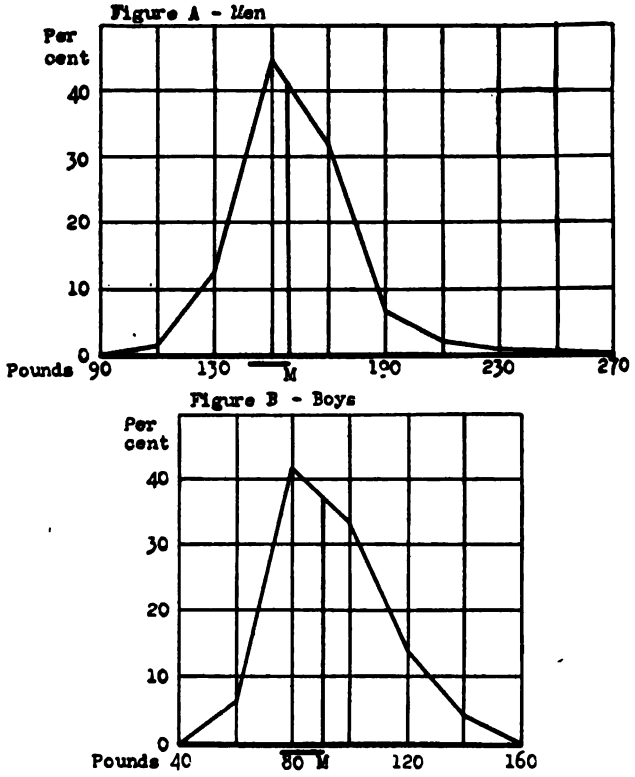
In Diagram 3 the scales for pounds are identical in both figures. The appearance of the diagram suggests that the two distributions are very much alike; as the figure for men has a greater spread at the base line than that for boys it would seem that the former represents, if anything, the wider dispersion. This impression is not borne out by the data. The

*The data, which are for boys attending the Worcester, Mass., public schools, are from a report by Franz Boas and Clark Wissler, published in the report of the U. S. Commissioner of Education for 1904.

Digitized by Google

actual dispersion (average deviation) is, roughly, the same for the two series: 14.2 pounds for the men and 14.3 pounds for the boys. But as the median for the men is 156.3 pounds, and that for the boys 90.8 pounds, computation shows that the significant measure of relative variability, the coefficient

DIAGRAM 3.—Weights of 738 men and 279 boys, shown without horizontal zeros.

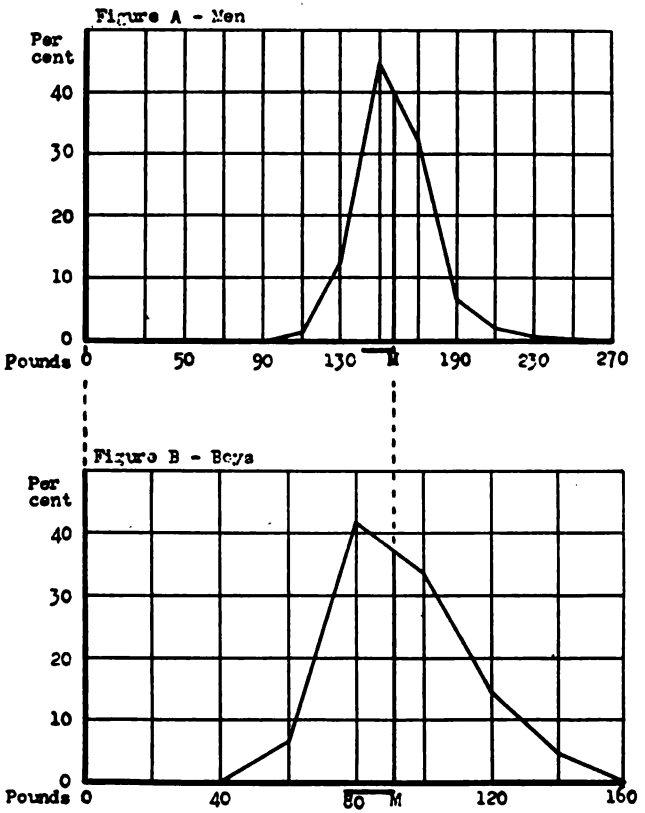


of dispersion, is .157 for the boys and only .091 for the men. In other words, the dispersion of the weights of the boys is 15.7 per cent. of the median weight of boys, while for the men the dispersion of the weights is but 9.1 per cent. of the median weight of men. The apparent similarity of the two distributions represented in Diagram 3 is, therefore, accidental and the diagram is misleading.

Digitized by Google

It may be said that anyone using Diagram 3 could determine the relative dispersions by a study of the figures of the scales; that the scales show the medians, and that it is not impossible to relate these medians to the dispersions. This is true, but, as the same facts can be determined from a fre-

DIAGRAM 4.—Weights of 738 men and 279 boys, shown with horizontal zeros.



quency table, the argument offered is merely an argument for not using graphical representations for comparing two or more series of variables.

Diagram 4 shows in graphic terms the true relationship between the dispersions. The base lines of Figures A and B of this diagram have been carried out to zero, and the scales

Digitized by Google

have been so adjusted that the distance from zero to the median is the same in both figures. It is now possible to view the dispersions in their relationship to the central tendencies. The lines representing the average deviations, as well as the contours of the curves, show very clearly that the weights of boys are much more widely dispersed than the weights of men.

The fact that in Diagram 4 the surface enclosed by the curve and base line of Figure B is much greater than that enclosed by the curve and base line of Figure A, might lead an incautious observer to assume that the dissimilarity in the appearance of the figures is due to a difference in the number of observations—that the number of boys exceeds the number of men. Such an inference would be unwarranted. As numbers have been reduced to percentages, 100 per cent. is the total for each group. The values are plotted upon the ordinates; hence, the spaces between the ordinates, and the areas enclosed by the curves and the base lines, are without significance. It is believed that the diagram affords a correct interpretation of the data; that it gives an impression of two groups of which one is somewhat closely clustered about its central tendency, while the other is much more widely dispersed.

It should be noted that there is an important group of frequency diagrams to which the arguments in favor of including the horizontal zero, which have been stated in the preceding pages, do not apply. These are diagrams of distributions in which the zero cannot be exactly located. In the so-called normal frequency distribution the base line and the ends of the curve are in asymptote—the ends and the base line are tangent at infinity. It follows that, in plotting probabilities, or results in the psychological field which are based not upon concrete measurements but upon rankings, the horizontal zero can not be shown.

But it is also impossible to show a zero based upon data of this kind in any type of diagram, and this is true whether the zero is vertical or horizontal. If the horizontal zero can not be shown in a frequency diagram representing the distribution of school boys with reference to a given mental trait, as determined by the rankings of competent judges, neither can a

Digitized by Google

zero be shown in a diagram in which the ability of any one of these boys at successive tests is indicated by a historical curve. It is possible to present a horizontal zero in a frequency diagram for any data for which a vertical zero for an ogive curve can be shown.

A practical objection to the inclusion of the horizontal zero is the fact that additional space is required. But this objection is no more applicable to the horizontal zero in frequency diagrams than to the vertical zero in line diagrams. The inclusion of the vertical zero in diagrams of the latter type is the established practice. And an inspection of the diagrams presented with this paper makes it clear that the inclusion of the horizontal zero presents no serious difficulties. A case will occasionally be encountered in which the dispersion constitutes so small a proportion of the central tendency that the zero, whether horizontal or vertical, must be omitted, but such cases are most exceptional.

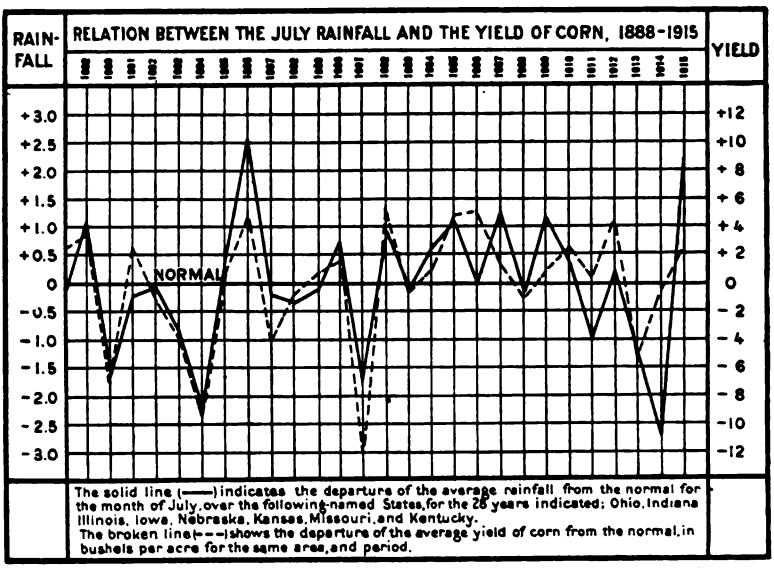
The arguments and the illustrations presented in the preceding pages seem to support the following conclusions: In frequency diagrams, where the position of the horizontal zero is exactly ascertainable, and where the dispersion is not too small in proportion to the measure of central tendency, the horizontal zero should be included in the diagram. This means that the horizontal zero should be included in a frequency diagram in all cases in which a zero for similar data would be included in any type of diagram. Without the horizontal zero the frequency diagram does not afford a complete graphical representation of the central tendency nor of the relationship of the central tendency to the distribution.

Digitized by Google

THE COEFFICIENT OF CORRELATION.*

By William Gardner Reed, *U. S. Weather Bureau.*

In many studies it is necessary or at least desirable to test the existence of concomitant variation between two series of variable quantities. A comparison of the plotted variables furnishes a rough, but for some purposes adequate, means of examining the relationship. Figure 1 is an example of this



U.S. Weather Bur. National Weather and Crop Bul, Ser. 1916 No. 14, June 20, 1916

sort of comparison. However, the use of curves is not to be recommended for careful work because of the difficulty in selecting the proper scales and the dangers resulting from personal bias. The usual tabular method is slightly more refined, but tables involve too many figures to give an adequate idea of the conditions and give no concise measure of the degree of relationship.

* For instructions for obtaining the coefficient of correlation see Persons, W. M.: The Correlation of Economic Statistics, Am. Stat. Asso., Quart. Publ., Vol. 12, pp. 287-322, Boston, 1910.

Digitized by Google

The English biometricians have perfected a method of stating the degree of relationship, which was invented by Bravais about 1845. "Correlation may be briefly defined as the tendency towards concomitant variation and the so-called correlation coefficient is simply a measure of such tendency, *more or less adequate according to the circumstances of the case.*"* The early statements of the use of the coefficient of correlation indicate clearly that the attempt to obtain such a coefficient from miscellaneous material is an abuse of this method of measuring relationship.† The material in hand should be investigated carefully before any attempt is made to determine the relationship by the use of the coefficient of correlation. This investigation may take the form of a correlation table or of a "dot chart" after Galton's graphic method of correlation.‡

METHOD OF PROCEDURE.

If the coefficient of correlation is to have any definite meaning, the procedure must be somewhat as follows:

1. The material (*e. g.* Table I) should be arranged in groups in the form of a correlation table (Table II), or, better, plotted as a dot chart (Figure 2). The table or chart should then be carefully examined to see whether the points may be generalized to a straight line, that is, whether there is a tendency for a high value of one variable to be associated with high values of the other variable and proportionately higher or lower values of the one to be associated with similar values of the other. This shows positive linear correlation. When lower values of the one are associated with higher values of the other, the correlation is said to be negative. For example, the dots in Figure 2 may be generalized to the line AB as well as to any curve.

* Brown, W.: The Essentials of Mental Measurement, Cambridge, University Press, 1911, p. 42. (Italics are the present writer's.)

† Yule, G. U.: Introduction to the Theory of Statistics, ed. 2, London, Griffin & Co., 1912, pp. 169, 177.

‡ See Davenport, C. B.: Statistical Methods, ed. 3, New York, Wiley, 1914, pp. 42–47.

Digitized by Google

TABLE I.

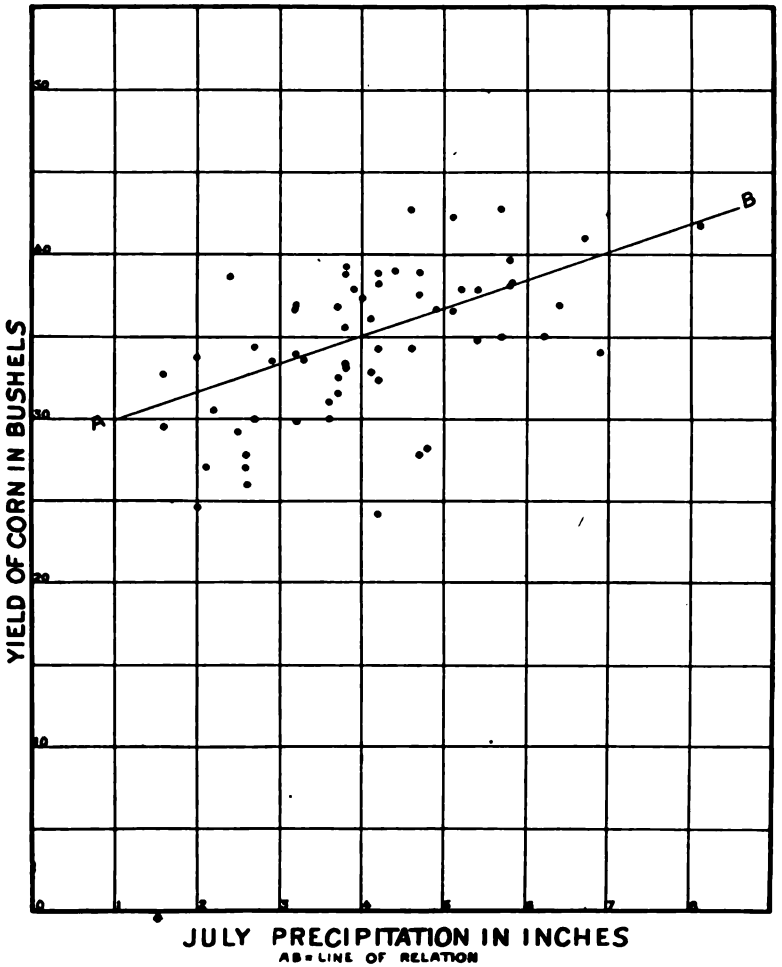
CORRELATION OF JULY RAINFALL AND THE YIELD OF CORN IN OHIO.

(Smith, J. W.: The Effect of Weather upon the Yield of Corn in Ohio. Washington, Mo., Weather Rev., Vol. 42, 1914, p. 80.

	July Rainfall.			Yield of Corn.			
Year.	Amount.	x.	x^2.	Bushels per Acre.	y.	y^2.	xy.
1854	2.6	− 1.4	1.96	26.0	− 9	81	+ 12.6
1855	5.8	+ 1.8	3.24	39.7	+ 5	25	+ 9.0
1856	2.6	− 1.4	1.96	27.7	− 7	49	+ 9.8
1857	4.9	+ .9	.81	36.6	+ 2	4	+ 1.8
1858	4.7	+ .7	.49	27.7	− 7	49	− 4.9
1859	1.6	− 2.4	5.76	29.5	− 5	25	+ 12.0
1860	5.8	+ 1.8	3.24	38.2	+ 3	9	+ 5.4
1861	3.3	− .7	.49	33.5	− 1	1	+ .7
1862	3.6	− .4	.16	30.0	− 5	25	+ 2.0
1863	2.6	− 1.4	1.96	27.0	− 8	64	+ 11.2
1864	2.1	− 1.9	3.61	27.0	− 8	64	+ 15.2
1865	5.7	+ 1.7	2.89	35.0	0	...	.....
1866	5.1	+ 1.1	1.21	36.5	+ 2	4	+ 2.2
1867	3.2	− .8	.64	29.8	− 5	25	+ 4.0
1868	2.7	− 1.3	1.69	34.4	− 1	1	+ 1.3
1869	4.8	+ .8	.64	28.4	− 7	49	− 5.6
1870	4.7	+ .7	.49	37.5	+ 3	9	+ 2.1
1871	3.7	− .3	.09	36.7	+ 2	4	− .6
1872	6.7	+ 2.7	7.29	40.9	+ 6	36	+ 16.2
1873	6.2	+ 2.2	4.84	35.1	0	....	.....
1874	3.8	− .2	.04	39.2	+ 4	16	− .8
1875	6.9	+ 2.9	8.41	34.2	− 1	1	− 2.9
1876	6.4	+ 2.4	5.76	36.9	+ 2	4	+ 4.8
1877	3.7	− .3	.09	32.5	− 2	4	+ .6
1878	5.4	+ 1.4	1.96	37.8	+ 3	9	+ 4.2
1879	4.2	+ .2	.04	34.3	− 1	1	− .2
1880	4.2	+ .2	.04	38.9	+ 4	16	+ .8
1881	3.6	− .4	.16	31.0	− 4	16	− 1.6
1882	3.2	− .8	.64	34.0	− 1	1	+ .8
1883	4.2	+ .2	.04	24.2	− 11	121	− 2.2
1884	3.8	− .2	.04	33.3	− 2	4	+ .4
1885	3.2	− .8	.64	36.8	+ 2	4	− 1.6
1886	2.9	− 1.1	1.21	33.5	− 1	1	+ 1.1
1887	2.2	− 1.8	3.24	30.5	− 4	16	+ 7.2
1888	4.4	+ .4	.16	38.9	+ 4	16	+ 1.6
1889	4.2	+ .2	.04	32.3	− 3	9	− .6
1890	2.0	− 2.0	4.00	24.6	− 10	100	+ 20.0
1891	3.8	− .2	.04	35.6	+ 1	1	− .2
1892	3.8	− .2	.04	33.3	− 2	4	+ .4
1893	2.5	− 1.5	2.25	29.1	− 6	36	+ 9.0
1894	1.6	− 2.4	5.76	32.6	− 2	4	+ 4.8
1895	2.0	− 2.0	4.00	33.7	− 1	1	+ 2.0
1896	8.1	+ 4.1	16.81	41.7	+ 7	49	+ 28.7
1897	4.6	+ .6	.36	34.3	− 1	1	− .6
1898	4.0	0	....	37.4	+ 2	4	.....
1899	4.2	+ .2	.04	38.1	+ 3	9	+ .6
1900	4.6	+ .6	.36	42.6	+ 8	64	+ 4.8
1901	2.7	− 1.3	1.69	30.0	− 5	25	+ 6.5
1902	4.7	+ .7	.49	38.8	+ 4	16	+ 2.8
1903	3.7	− .3	.09	31.5	− 3	9	+ .9
1904	4.1	+ .1	.01	32.8	− 2	4	− .2
1905	3.9	− .1	.01	37.9	+ 3	9	− .3
1906	5.1	+ 1.1	1.21	42.2	+ 7	49	+ 7.7
1907	5.4	+ 1.4	1.96	34.8	0	....	.....
1908	4.1	+ .1	.01	36.1	+ 1	1	+ .1
1909	3.8	− .2	.04	38.7	+ 4	16	− .8
1910	3.2	− .8	.64	36.6	+ 2	4	− 1.6
1911	2.4	− 1.6	2.56	38.6	+ 4	16	− 6.4
1912	5.7	+ 1.7	2.89	42.8	+ 8	64	+ 13.6
1913	5.2	+ 1.2	1.44	37.8	+ 3	9	+ 3.6
		−30.2	112.67		−125	1258	+201.4
		+34.1			+ 99		
		+ 3.9			− 26		

Digitized by Google

CORRELATION BETWEEN JULY PRECIPITATION AND YIELD OF CORN IN OHIO



Digitized by Google

$M'_x = 4.0$ inches

$M_x = 4.0 + \frac{3.9}{60} = 4.1$

$\Sigma x = +3.9$

$\Sigma x^2 = 112.67$

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$= \sqrt{\frac{12.67}{60} - .0036}$$

$= 1.4$

$M'_y = 35$ bu.

$M_y = 35 - \frac{26}{60} = 34.6$

$\Sigma y = -26$

$\Sigma y^2 = 1258$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2}$$

$$= \sqrt{\frac{1258}{60} - .2}$$

$= 4.6$

$\Sigma xy = 201.4$

$$r = \frac{\frac{\Sigma xy}{n} - \left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\sigma_x \sigma_y}$$

$$= \frac{\frac{201.4}{60} - \frac{3.9}{60} \times \frac{-26}{60}}{1.4 \times 4.6}$$

$$= \frac{3.36 + .03}{6.44}$$

$= 0.526 \pm E_r$

$$E_r = \pm .674 \frac{1 - r^2}{\sqrt{n}}$$

$$= \pm .674 \frac{.723}{7.7}$$

$= \pm .063$

$r = +0.526 \pm .063$

NOTE: r is not the same here as in the original paper because a single average yield of corn has been used for simplicity.

EXPLANATION OF SYMBOLS.

n number of observations (years of record)
M_x true mean July precipitation
M'_x some arbitrary number near M_x
M_y true mean yield of corn
M'_y some arbitrary number near M_y
x departure of each July precipitation from M'_x
y departure of yield of corn in each year from M'_y

Digitized by Google

Σx Algebraic sum of departures of July precipitation

Σy Algebraic sum of departures of yield of corn

Σx^2 algebraic sum of squares of departures of July precipitation

Σy^2 algebraic sum of squares of departures of yield of corn

Σxy algebraic sum of products of departures (x and y)

σ_x standard deviation of July precipitation

$$\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

σ_y standard deviation of yield of corn

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2}$$

r coefficient of correlation

$$r = \frac{\frac{\Sigma xy}{n} - \left(\frac{\Sigma x}{n}\right)\left(\frac{\Sigma y}{n}\right)}{\sigma_x \sigma_y}$$

E_r probable error of the coefficient of correlation

$$E_r = \pm .674 \frac{1 - r^2}{\sqrt{n}}$$

TABLE II.

CORRELATION TABLE SHOWING THE RELATION BETWEEN JULY PRECIPITATION AND THE YIELD OF CORN IN OHIO.

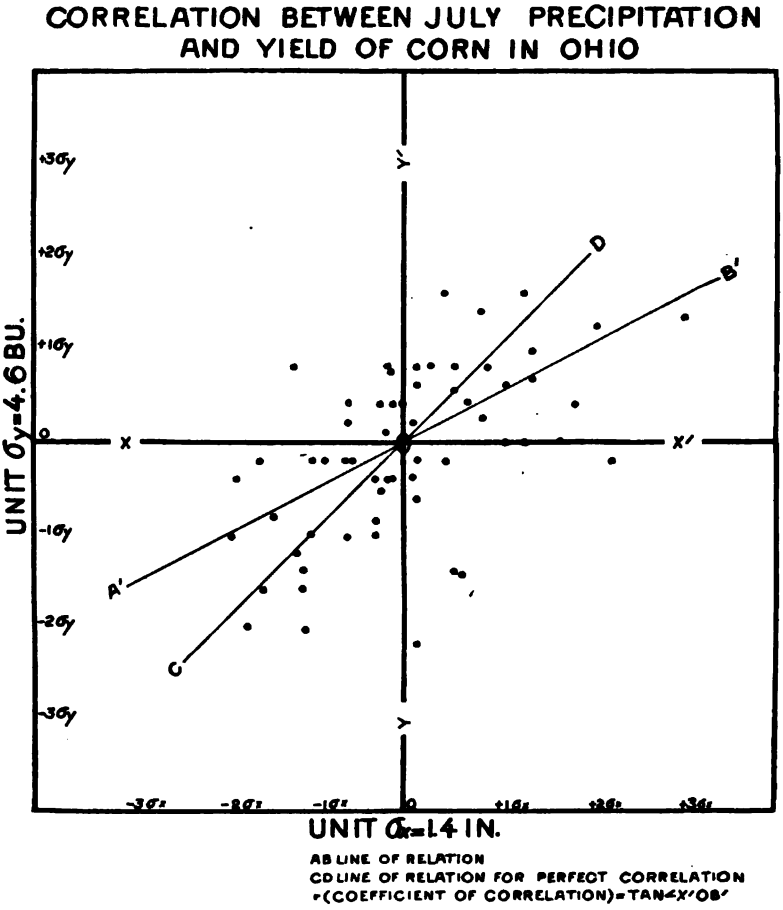
(From Smith, J. W.: The Effect of Weather on the Yield of Corn, Washington, Mo., Weather Rev., Vol. 42, 1914, pp. 78–93.)

YIELD OF CORN IN BUSHELS PER ACRE.

July Precipitation in Inches.	20.0 to 24.9	25.0 to 29.9	30.0 to 34.9	35.0 to 39.9	40.0 to 44.9
80–89	—	—	—	—	1
70–79	—	—	—	—	—
60–69	—	—	1	1	1
50–59	—	—	1	7	2
40–49	1	2	4	8	1
30–39	—	1	8	7	—
20–29	—	5	5	1	1
10–19	—	1	1	—	—

Digitized by Google

2. If it appears from this examination that a straight line is as good a fit as any other type of curve not too complicated to be useful as a measure of relationship, the data may be



replotted on a new dot chart for which the unit of measurement on one axis is the standard deviation of one of the variables, and the unit on the other axis is the standard deviation of the other variable (see Figure 3).

Digitized by Google

3. The position of the straight line which most nearly satisfies the data on the second dot chart may be determined rigidly by the method of least squares. When the standard deviation of one variable is used as the unit of the ordinates and the standard deviation of the other variable as the unit of the abscissae, the angles between this straight line of closest fit and the axes are significant. If these angles are equal, *i. e.* each 45°, the relationship between the variables is perfect (see C–D in Figure 3). If the line coincides with one axis or the other no relationship is shown, although the converse is not necessarily true.* Positions between these two show partial relationship (see A′B′ in Figure 3).

4. The coefficient of correlation is merely a statement of the position of the straight line of closest fit on a chart where the units are the standard deviations of the variables as this position is determined by the least square adjustment.† The coefficient of correlation is expressed as the tangent of the angle made by the line of closest fit and the axis to which it is more nearly parallel (*e. g.* angle X′OB′ in Figure 3 is 27½°, tan X′OB′ = +0.526). In actual practice the coefficient of correlation may be determined mathematically from the data as shown in Table I without plotting the material on a dot chart, like Figure 3. However, the coefficient should never be attempted without first investigating the relationship far enough to see if it follows a straight line. That is, steps 2 and 3 may be omitted in practice; step 1 should never be omitted.

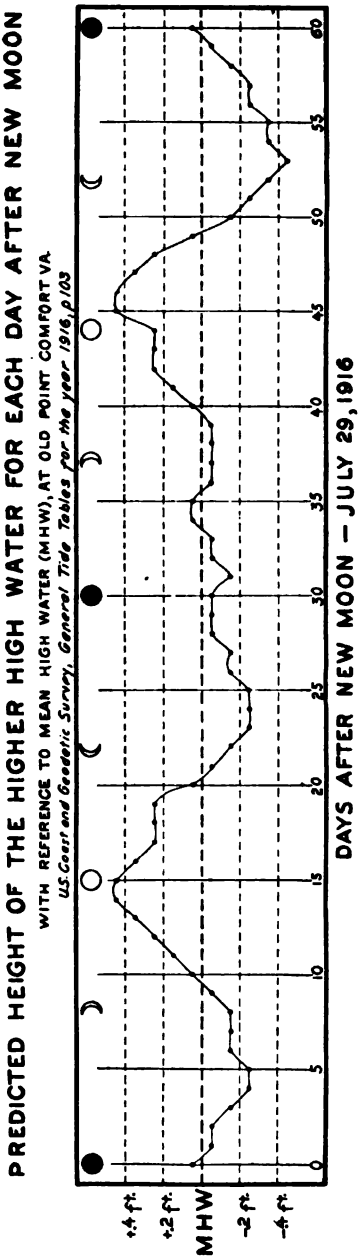
5. If the examination of the correlation table or dot chart shows that the relation is not that of a simple straight line, the coefficient of correlation is not a measure of the relationship between the variables.

LIMITATIONS OF THE COEFFICIENT OF CORRELATION.

It is clear even from a superficial study of the question that the coefficient of correlation obtained from material where a straight line relationship does not obtain may be too small,

* Yule, G. U.: Introduction to the Theory of Statistics, ed. 2, London, Griffin & Co., 1912, pp. 174–175.
† Yule, G. U.: Introduction to the Theory of Statistics, ed. 2, London, Griffin & Co., 1912, p. 172.

Digitized by Google



but will never be too large.* A coefficient of correlation may be near zero when there is very close relationship, as is shown in such a condition as the relationship between the height of high water and the phase of the moon which is shown for Old Point Comfort, Va., by Table III and Figure 4. The figure indicates that the relation is harmonic; although there is a close and very definite relation between the phenomena, the coefficient of correlation is near zero ($-0.106 \pm .088$) because the different portions of the curve of regression are in such relations to each other that a straight line along an axis will most nearly satisfy all the points. Of course the angle is then zero and its tangent is zero.

*See Yule, G. U.: Introduction to the Theory of Statistics, ed. 2, London, Griffin & Co., 1912, p. 175, and Brown, W.: The Essentials of Mental Measurement, Cambridge, University Press, 1911, p. 27–59.

Digitized by Google

TABLE III.

CORRELATION OF TIME AFTER NEW MOON AND PREDICTED HEIGHT OF THE HIGHER HIGH WATER AT OLD POINT COMFORT, VA.

(U. S. Coast and Geodetic Survey, General tide tables for the year, 1916, p. 103.)

Days after New Moon July 29, 1916.	x.	x^2.	Height above M. L. W.	y.	y^2.	xy.
0	−30	900	2.7	+ .1	.01	− 3.
1	−29	841	2.6	0		
2	−28	784	2.6	0		
3	−27	729	2.5	− .1	.01	+ 2.7
4	−26	676	2.4	− .2	.04	+ 5.2
5	−25	625	2.4	− .2	.04	+ 5.0
6	−24	576	2.5	− .1	.01	+ 2.4
7	−23	529	2.5	− .1	.01	+ 2.3
8	−22	484	2.5	− .1	.01	+ 2.2
9	−21	441	2.6	0		
10	−20	400	2.7	+ .1	.01	− 2.0
11	−19	361	2.8	+ .2	.04	− 3.8
12	−18	324	2.9	+ .3	.09	− 5.4
13	−17	289	3.0	+ .4	.16	− 6.8
14	−16	256	3.1	+ .5	.25	− 8.0
15	−15	225	3.1	+ .5	.25	− 7.5
16	−14	196	3.0	+ .4	.16	− 5.6
17	−13	169	2.9	+ .3	.09	− 3.9
18	−12	144	2.9	+ .3	.09	− 3.6
19	−11	121	2.9	+ .3	.09	− 3.3
20	−10	100	2.7	+ .1	.01	− 1.0
21	− 9	81	2.6	0		
22	− 8	64	2.5	− .1	.01	+ .8
23	− 7	49	2.4	− .2	.04	+ 1.4
24	− 6	36	2.4	− .2	.04	+ 1.2
25	− 5	25	2.4	− .2	.04	+ 1.0
26	− 4	16	2.5	− .1	.01	+ .4
27	− 3	9	2.5	− .1	.01	+ .3
28	− 2	4	2.6	0		
29	− 1	1	2.6	0		
30	0	0	2.6	0		
31	+ 1	1	2.5	− .1	.01	− .1
32	+ 2	4	2.6	0		
33	+ 3	9	2.6	0		
34	+ 4	16	2.7	+ .1	.01	+ .4
35	+ 5	25	2.7	+ .1	.01	+ .5
36	+ 6	36	2.6	0		
37	+ 7	49	2.6	0		
38	+ 8	64	2.6	0		
39	+ 9	81	2.6	0		
40	+10	100	2.7	+ .1	.01	+ 1.0
41	+11	121	2.8	+ .2	.04	+ 2.2
42	+12	144	2.9	+ .3	.09	+ 3.6
43	+13	169	2.9	+ .3	.09	+ 3.9
44	+14	196	2.9	+ .3	.09	+ 4.2
45	+15	225	3.1	+ .5	.25	+ 7.5
46	+16	256	3.1	+ .5	.25	+ 8.0
47	+17	289	3.0	+ .4	.16	+ 6.8
48	+18	324	2.9	+ .3	.09	+ 5.4
49	+19	361	2.7	+ .1	.01	+ 1.9
50	+20	400	2.5	− .1	.01	− 2.0
51	+21	441	2.4	− .2	.04	− 4.2
52	+22	484	2.3	− .3	.09	− 6.6
53	+23	529	2.2	− .3	.16	− 9.2
54	+24	576	2.3	− .3	.09	− 7.2
55	+25	625	2.3	− .3	.09	− 7.5
56	+26	676	2.4	− .2	.04	− 5.2
57	+27	729	2.4	− .2	.04	− 5.4
58	+28	784	2.5	− .1	.01	− 2.8
59	+29	841	2.6	0		
60	+30	900	2.8	+ .2	.04	+ 6.0
		18910		−3.9 +6.9	3.24	−25.1
				+3.0		

Digitized by Google

$$M'_x = 30 \qquad M'_y = 2.6$$

$$M_x = 30 \qquad M_y = 2.6 + \frac{3.0}{61} = 2.65$$

$$\Sigma x = 0 \qquad \Sigma y = +.05$$

$$\Sigma x^2 = 18910 \qquad \Sigma y^2 = 3.24$$

$$\sigma_x = \sqrt{\frac{18910}{61} - 0} \qquad \sigma_y = \sqrt{\frac{3.24}{61} - .003}$$

$$= 17.6 \qquad = .22$$

$$\Sigma xy = -25.1$$

$$r = \frac{\frac{-25.1}{61} - 0}{17.6 \times .22}$$

$$= -\frac{.411}{3.87}$$

$$= -.106 \pm E_r$$

$$E_r = .674\,\frac{1-(-.106)^2}{\sqrt{61}}$$

$$= .674\,\frac{1-.0112}{7.8}$$

$$= 0.674 \times 0.13$$

$$r = -0.106 \pm 0.088$$

When the relation is not linear the concomitant variation may be shown by the use of a "correlation ratio," which is simply a further development of the theory of correlation.*

It is, however, not the purpose of this paper to consider relationships shown by curves of a higher order than a straight line, as such correlations involve more complicated mathematical theory and also require many more observations to be significant.

ADEQUACY OF THE COEFFICIENT OF CORRELATION.

The conclusion seems legitimate that the coefficient of correlation may be used strictly as a *measure* of relationship,

* See Pearson, K.: Mathematical Contributions to the Theory of Evolution: 14. On the general theory of skew correlation and non-linear regression. London, Drapers Company Research Memoirs. Biometric Series 2, 1905. Brown, W.: The Essentials of Mental Measurement. Cambridge, University Press, 1911, pp. 57–59.

Digitized by Google

when such relationship has been determined by other investigation to follow straight line relations. The use of the coefficient of correlation is to be recommended because it is independent of the personal equation of the investigator, and of the units employed, and because it shows rigidly the correct position of the line indicated by the dot chart.

In using the coefficient of correlation it is desirable to calculate the probable error (see Tables I and III for method).* The probable error is that divergence from the observed mean on either side within which half the observations lie. Its size is a measure of how closely the results from an infinite number of cases would correspond with those obtained from the observed cases. When the coefficient of correlation is not greater than its probable error there is no evidence that there is any correlation; but when the coefficient of correlation is clearly greater than its probable error correlation is indicated; and when it is much greater (six times as great is an accepted empirical amount) it may be safely assumed that there is concomitant variation.†

The coefficient of correlation is obtained by applying the least square adjustment to all the material and is, therefore, the straight line of closest fit. If the relationship is not that of a straight line, it is obvious that the straight line of closest fit is not a good measure of the relationship and that some other measure (*e. g.*, the correlation ratio) must be used. Therefore, the coefficient of correlation should never be used to show relationship until after the phenomena have been investigated, at least far enough to show whether a straight line satisfies the relationship as well as any other curve.

LITERATURE.

The development of the theory of correlation resulting in the adoption and use of the coefficient of correlation is, of course, largely mathematical. While the literature on the subject is considerable, the greater part of the contributions are concerned with the application of the coefficient to par-

* For a general discussion of the significance of probable error see Yule, G. U. Introduction to the Theory of Statistics, ed. 2, London, Griffin & Co., 1912, pp. 310–311.

† See Bowley, A. L.: Elements of Statistics, ed. 3, New York, Scribner, 1907, p. 320.

Digitized by Google

ticular problems, and hence the development of the theory of correlation is incidental and widely scattered.

"The fundamental theorems of correlation were for the first time and almost exhaustively discussed by A. Bravais* . . . [more than] half a century ago. He deals completely with the correlation of two and three variables. Forty years later Mr. J. D. Hamilton Dickson† dealt with a special problem proposed to him by Mr. Galton, and reached on a somewhat narrow basis some of Bravais' results for correlation of two variables. Mr. Galton at the same time introduced an improved notation which may be summed up in the 'Galton function' or coefficient of correlation. This indeed appears in Bravais' work, but a single symbol is not used for it. In 1892 Professor Edgeworth, also unconscious of Bravais' memoir, dealt in a paper on 'Correlated Averages' with correlation for three variables.‡ He obtained results identical with Bravais, although expressed in terms of 'Galton's functions.' "§

The following publications contain complete statements of the later development:

PEARSON, KARL: Contributions to the mathematical theory of evolution; London, Royal Society, Philosophical Transactions, Series A, as follows:

1. On the dissection of frequency curves, Vol. 185, 1894, pp. 71–110.
2. Skew variations in homogeneous material, Vol. 186, 1895, pp. 343–414.
3. Regression, heredity, and panmixia, Vol. 187, 1896, pp. 253–318.
4. On the probable errors of frequency constants and on the influence of random selection on variation and correlation, Vol. 191, 1898, pp. 229–311.
5. On the reconstruction of the stature of prehistoric races, Vol. 192, 1898, pp. 169–244.
6. Genetic (reproductive) selection; inheritance of fertility in man and of fecundity in thoroughbred race horses, Vol. 192, 1899, pp. 257–330.
7. On the correlation of characters not quantitatively measureable, Vol. 195, 1900, pp. 1–47.

* *Analyse mathematique sur les probabilities des erreurs de situation d'un point.* Paris, Academie des Sciences, *Memoires presentes par divers savants.* Series 2, Vol. 9, 1846, pp. 255–332.

† Appendix to Galton, F.: Family Likeness in Stature. London, Royal Society, Proceedings, Vol. 40, 1886, pp. 63–73.

‡ London, Philosophical Magazine, Series 5, Vol. 34, 1892, pp. 190–204.

§ Pearson, Karl: London Royal Society Philosophical Transactions, Series A, Vol. 187, 1896, p. 261.

Digitized by Google

8. On the inheritance of characters not quantitatively measureable, Vol. 195, 1900, pp. 75–150.
9. On the principle of homotyposis and its relation to heredity, to the variability of the individual, and to that of the race, Vol. 197, 1901, pp. 285–379.
10. Supplement to a memoir on skew variation, Vol. 197, 1901, pp. 443–459.
11. On the influence of natural selection on the variability and correlation of organs, Vol. 200, 1902, pp. 1–66.
12. On a generalized theory of alternative inheritance with special reference to Mendel's Laws, Vol. 203, 1904, pp. 53–86.

In London, Drapers' Company Research Memoirs, Biometric Series.

13. On the theory of contingency and its relation to association and normal correlation. Memoir 1.
14. On the general theory of skew correlation and non-linear regression. Memoir 2.
15. On the mathematical theory of random migration. Memoir 3, 1906.
16. On further methods of determining correlation. Memoir 4, 1907.
17. [Not published.]
18. On a novel method of regarding the association of two variates classed solely in alternate categories. Memoir 7, 1912.

Pearson, Karl: On the partial correlation ratio. London, Royal Society, Proceedings, Series A, Vol. 91, 1915, pp. 492–498.

Brown, W.: The essentials of mental measurement, Cambridge, University Press, 1911.

Elderton, W. P.: Frequency curves and correlation. London, Layton Brothers, 1906.

Hooker, R. H.: Correlation of successive observations, Royal Statistical Society Journal, Vol. 68, pp. 676–703.

Tolley, H. R.: The theory of correlation as applied to farm survey data on fattening baby beef, U. S. Department of Agriculture Bul. 504, Washington, Govt. Ptg. Office, 1917.

Walker, Gilbert T.: Correlation in seasonal variation of weather, Indian Meteorological Department Memoirs, Simla, 1909–1915.

1. Correlation in season variation of climate, Vol. 20, part 6, 1909, pp. 117–124.
2. (A) On the probable error of a coefficient of correlation with a group of factors.
   (B) Some applications of statistical methods to seasonal forecasting, Vol. 21, part 2, 1910, pp. 22–45.
3. On the criterion for the reality of relationships or periodicities, Vol. 21, part 9, 1914, pp. 13–16.
4. Sunspots and rainfall, Vol. 21, part 10, 1915, pp. 17–60.
5. Sunspots and temperature, Vol. 21, part 11, 1915, pp. 61–90.
6. Sunspots and pressure, Vol. 21, part 12, 1915, pp. 91–118.

Digitized by Google

YULE, G. UDNY: Introduction to the theory of statistics, ed. 2, London, C. Griffin & Co., 1912, pp. 157–253.

More elementary discussions are contained in the following papers:

PERSONS, W. M.: The correlation of economic statistics. Boston, American Statistical Association, Quarterly Publications, Vol. 12 (1910), pp. 287–322.

HOOKER, R. H.: An elementary explanation of correlation: illustrated by rainfall and the depth of water in a well; London, Royal Meteorological Society Quarterly Journal, Vol. 34, 1908, pp. 277–291.

ELDERTON, W. P. and E. M.: Primer of statistics, London, A. and C. Black, 1910, pp. 55–72.

KING, W. I.: Elements of statistical method, New York, Macmillan, 1912, pp. 197–215.

DINES, W. H.: The practical application of statistical methods to meteorology. London, H. M. Meteorological Office, The computer's handbook (M. O. 223), section 5, part 2, 1915, pp. V29–V52.

The most complete bibliographies will be found in:

YULE, G. UDNY: Introduction to the theory of statistics, London, C. Griffin & Co., 1912, pp. 188, 208–209, 225–226, and 252.

DAVENPORT, C. B.: Statistical methods with special reference to biological variation, third, revised edition, New York, J. Wiley & Sons, 1914, pp. 62 and 85–104.

Digitized by Google

AN INTERPOLATION FORMULA FOR "EQUIDISTANT" FREQUENCY DISTRIBUTIONS.

By Harry Langman, *Ocean Accident & Guarantee Corporation, New York City.*

A frequency distribution is conveniently represented by the area under some continuous curve. Moreover, the total frequency is usually included between two distinct values of the independent variable, outside of which there is no frequency. Hence, if x denote the independent variable, and $y=f(x)$ denote the curve bounding the distribution, we must have

$$f(a)=f(b)=0, \tag{1}$$

where a and b are the extreme values of x.

The frequency between any two values of x, α and β, is given by $\int_{\alpha}^{\beta} y\,dx$. In the usual case the values of this integral are known for each of the sub-intervals obtained by dividing the entire internal $[a, b]$ into equal parts. Let m be the number of sub-intervals, and for convenience choose each sub-interval of unit length. Then

$$b=a+m. \tag{2}$$

Let a_k, where $k=1, 2, 3, \cdots, m$, be the known frequencies over each of the sub-intervals $[a, a+1]$, $[a+1, a+2]$, $\cdots$, $[a+k-1, a+k]$, $\cdots$, $[a+m-1, a+m]$. Then

$$a_k=\int_{a+k-1}^{a+k} y\,dx;\ k=1, 2, 3, \cdots, m. \tag{3}$$

The problem then is to interpolate frequencies for other than the equidistant values of the independent variable. This consists in finding some function $f(x)$ which will satisfy the $m+2$ conditions of (1) and (3).* Since this is infinitely

*Cf. the problem first solved by Lagrange, Oeuvres 1, p. 87. Also articles in Ensyk. d. Math. Wissen. IIA, 9a, p. 647.

Digitized by Google

indeterminate, the problem reduces to finding a form for $f(x)$ which can be handled practically, and which can be applied in all cases.

For this purpose we may choose for $y=f(x)$ the function

$$y=\sum_{n=1}^{n=m}\frac{n\pi}{m}\cdot a_n\cdot sin\,\frac{(x-a)\,n\pi}{m}. \tag{4}$$

By (2), this evidently satisfies the conditions (1). The conditions (3) then become

$$a_k=\sum_{n=1}^{n=m}a_n\cdot\left[\cos\frac{(k-1)\,n\pi}{m}-\cos\frac{kn\pi}{m}\right];\ k=1,2,3,\cdots,m. \tag{5}$$

The m conditions (5) are just sufficient to determine the m constants a_n. The problem now is to evaluate these coefficients.

Suppose the m equations of (5) to be written in succession and that we add the first k of them together, for all values of k. Then, if we put

$$s_k=a_1+a_2+a_3+\cdots+a_k, \tag{6}$$

$$s_k=\sum_{n=1}^{n=m}a_n\cdot\left[1-\cos\frac{n\pi}{m}k\right]=\sum_{n=1}^{n=m}a_n-\sum_{n=1}^{n=m}a_n\cdot\cos\frac{n\pi}{m}k.$$

Letting

$$A=\sum_{n=1}^{n=m}a_n, \tag{7}$$

the last may be written

$$s_k=A-\sum_{n=1}^{n=m}a_n\cos\frac{n\pi}{m}k;\ k=1,2,3,\cdots,m. \tag{8}$$

The m conditions of (8) replace those of (5).

Let r be any integer from 1 to m. Then

$$r\leqq m;\ n\leqq m. \tag{9}$$

Now consider the m quantities

$$2\cos\frac{r\pi}{m}k;\ k=1,2,3,\cdots,m. \tag{10}$$

Multiplying each of the equations of (8) by the corresponding quantity of (10) and adding the resulting equations we obtain

$$2\sum_{k=1}^{k=m}s_k\cos\frac{r\pi}{m}k=A\cdot I\,(r)+\sum_{n=1}^{n=m}a_n\cdot K\,(n,r), \tag{11}$$

Digitized by Google

where

$$I(r) = 2 \sum_{k=1}^{k=m} \cos \frac{r\pi}{m} k, \tag{12}$$

and

$$K(n, r) = -\sum_{k=1}^{k=m} 2 \cos \frac{n\pi}{m} k \cdot \cos \frac{r\pi}{m} k. \tag{13}$$

From trigonometry, we have

$$\sum_{p=1}^{p=q'} \cos p\theta = -\frac{1}{2} + \frac{\sin\left(q\theta + \frac{\theta}{2}\right)}{2 \sin \frac{\theta}{2}}. \tag{14}$$

Putting $q = m$, and $\theta = \frac{r\pi}{m}$ in (14), (12) becomes

$$I(r) = -1 + \frac{\sin\left[r\pi + \frac{r\pi}{2m}\right]}{\sin \frac{r\pi}{2m}}. \tag{15}$$

By (9), the denominator of (15) never vanishes. Hence:

(16) *if r is even, $I(r) = 0$; if r is odd, $I(r) = -2$.*

From trigonometry,

$$2 \cos x \cdot \cos y = \cos(x+y) + \cos(x-y).$$

Applying this, (13) becomes

$$\begin{aligned} K(n, r) &= -\sum_{k=1}^{k=m} \left[\cos \frac{(n+r)\pi}{m} k + \cos \frac{(n-r)\pi}{m} k\right] \\ &= -\sum_{k=1}^{k=m} \cos \frac{(n+r)\pi}{m} k - \sum_{k=1}^{k=m} \cos \frac{(n-r)\pi}{m} k. \end{aligned} \tag{17}$$

Again using (14), (17) becomes

$$\begin{aligned} K(n, r) = \frac{1}{2} - \frac{\sin\left[(n+r)\pi + \frac{(n+r)\pi}{2m}\right]}{2 \sin \frac{(n+r)\pi}{2m}} \\ + \frac{1}{2} - \frac{\sin\left[(n-r)\pi + \frac{(n-r)\pi}{2m}\right]}{2 \sin \frac{(n-r)\pi}{m}}. \end{aligned} \tag{18}$$

Now suppose $n+r$ is *even*; then $n-r$ is even, and the angles

Digitized by Google

$(n+r)\pi$ and $(n-r)\pi$ may be dropped from the brackets in (18). In that case $K=0$, except when the denominator of either fraction vanishes. This can happen only when $\frac{n+r}{2m}$ or $\frac{n-r}{2m}$ is integral. By (9), $\frac{n+r}{2m}$ is integral only when $n=r=m$; $\frac{n-r}{2m}$ is integral only when $n=r$. Hence we have:

(19) *if* $n+r$ *is even, and* $n \neq r$, $K(n_1 r)=0$.

Now suppose $n+r$ is even and $n=r$. In this case (18) is indeterminate. However, on returning to (17) we obtain

$$(18a) \qquad K(r,r)=\frac{1}{2}-\frac{\sin\left[2r\pi+\frac{r\pi}{m}\right]}{2\sin\frac{r\pi}{m}}-m=\frac{1}{2}-\frac{\sin\frac{r\pi}{m}}{2\sin\frac{r\pi}{m}}-m.$$

Hence $K=-m$, unless the denominator is zero. This occurs only when $r=m$. Hence we have:

(20) *if* $r \neq m$, $K(r,r)=-m$.

Suppose $n=r=m$. Substituting in (17) we get at once:

(21) $$K(m,m)=-2m.$$

Consider $n+r$ *odd*. Then $n-r$ is odd, and the angles $(n+r)\pi$ and $(n-r)\pi$ in the brackets of (18) may each be replaced by π. It is clear, furthermore, that in this case the denominators of (18) cannot vanish. Hence we have

(22) *if* $n+r$ *is odd*, $K(n,r)=2$.

For definiteness, we shall now suppose m *odd*. Then we may conveniently represent the results of (16), (19), (20), (21), (22), in the following table:

r even	n even	$n=r$	$K=-m$
		$n \neq r$	$K=0$
	n odd		$K=2$

$I=0$

Digitized by Google

$$r \text{ odd} \begin{cases} n \text{ odd} \begin{cases} n=r \begin{cases} r=m & K=-2m \\ r \neq m & K=-m \end{cases} \\ n \neq r & K=0 \end{cases} \\ n \text{ even} & K=2 \end{cases} \Bigg\} I=-2.$$

We note that for each value of r from 1 to m we get a set of multipliers in (10), and a new equation (11) from equations (8). We thus obtain m new equations to replace those of (8). That the new set is equivalent to (8) is shown by the fact that a definite solution is obtained.

We may put

$$M = a_1 + a_3 + a_5 + \cdots + a_m; \tag{23}$$
$$N = a_2 + a_4 + a_6 + \cdots + a_{m-1}.$$

Then

$$A = M + N.$$

Consider first the irregular case $r=m$. From the table we obtain: $I=-2$; for n even, $K=2$; for n odd and $n \neq m$, $K=0$; for $n=r=m$, $K=-2m$. Hence equation (11) takes the form

$$2\sum_{k=1}^{k=m} s_k \cos \pi k = -2A - 2ma_m + 2(a_2 + a_4 + a_6 + \cdots + a_{m-1})$$
$$= -2A + 2N - 2ma_m,$$

or

$$\sum_{k=1}^{k=m} (-1)^k s_k = -M - \tfrac{1}{2}ma_m. \tag{24}$$

Let r assume any of the *even* values from 2 to $m-1$. From the table, $I=0$; $K=2$ if n is odd; $K=0$ if n is even and $n \neq r$; $K=-m$ if $n=r$. Hence (11) becomes

$$2\sum_{k=1}^{k=m} s_k \cdot \cos \frac{r\pi}{m} k = 2(a_1 + a_3 + \cdots + a_m) - ma_r$$
$$= 2M - ma_r; \; r = 2, 4, 6, \cdots, m-1. \tag{25}$$

Let r assume any of the *odd* values allowed by (9), exclusive of $r=m$, already considered. From the table we then obtain, $I=-2$; $K=2$ if n is even; $K=0$ if n is odd and $n \neq r$; $K=-m$ if $n=r$. Hence (11) becomes

Digitized by Google

$$(26)\quad 2\sum_{k=1}^{k=m} s_k \cdot \cos\frac{r\pi}{m}k = -2A+2(a_2+a_4+\cdots+a_{m-1})-ma_r$$
$$= -2M-ma_r;\ r=1,3,5,\cdots,m-2.$$

We have then m equations in (24), (25), (26). We now proceed to determine M. Adding together the $\frac{m-1}{2}$ equations of (26), we obtain

$$2\sum_{t=1}^{t=\frac{m-1}{2}}\sum_{k=1}^{k=m} s_k \cdot \cos\frac{(2t-1)\pi k}{m} = -(m-1)M-m(a_1+a_3$$
$$+a_5+\cdots+a_{m-2})$$
$$(27)\qquad = -(m-1)M-m(M-a_m)$$
$$= -(2m-1)M+ma_m.$$

Adding (24) and (27), we obtain

$$(28)\quad M = -\frac{1}{2m}\left\{\sum_{k=1}^{k=m}(-1)^k s_k+2\sum_{t=1}^{t=\frac{m-1}{2}}\sum_{k=1}^{k=m} s_k \cdot \cos\frac{(2t-1)\pi}{m}k\right\}.$$

This gives us M in terms of definitely known quantities. From (24), (25), (26), the coefficients a_n are then determined. We can, however, decidedly simplify the expression for M in (28).

From trigonometry,

$$\sum_{u=1}^{u=t}\cos(2u-1)\theta = \frac{\sin 2t\theta}{2\sin\theta},$$

unless $\sin\theta = 0$. Using this, (28) becomes in order

$$M = -\frac{1}{2m}\left\{\sum_{k=1}^{k=m}(-1)^k s_k+2\sum_{k=1}^{k=m} s_k \sum_{t=1}^{t=\frac{m-1}{2}}\cos(2t-1)\frac{\pi k}{m}\right\}$$

$$= -\frac{1}{2m}\left\{\sum_{k=1}^{k=m}(-1)^k s_k+2\sum_{k=1}^{k=m-1} s_k \sum_{t=1}^{t=\frac{m-1}{2}}\cos(2t-1)\frac{\pi k}{m}\right.$$
$$\left.+2\,s_m\sum_{t=1}^{t=\frac{m-1}{2}}\cos(2t-1)\pi\right\}$$

Digitized by Google

$$= -\frac{1}{2m}\left\{\sum_{k=1}^{k=m}(-1)^k s_k + 2\sum_{k=1}^{k=m-1} s_k \frac{\sin\frac{(m-1)\pi k}{m}}{2\sin\frac{\pi k}{m}} - (m-1)s_m\right\}$$

$$= -\frac{1}{2m}\left\{\sum_{k=1}^{k=m}(-1)^k s_k + \sum_{k=1}^{k=m-1} s_k \frac{\sin\pi k\cdot\cos\frac{\pi k}{m} - \cos\pi k\cdot\sin\frac{\pi k}{m}}{\sin\frac{\pi k}{m}} - (m-1)s_m\right\}$$

$$= -\frac{1}{2m}\left\{\sum_{k=1}^{k=m}(-1)^k s_k - \sum_{k=1}^{k=m-1}(-1)^k s_k - (m-1)s_m\right\}$$

$$= -\frac{1}{2m}\left\{-s_m - ms_m + s_m\right\}$$

(29) $$= \frac{s_m}{2}.^*$$

From (24), (25), (26), we now obtain the coefficients a_n in the following forms:

$$\left.\begin{aligned} a_m &= -\frac{1}{m}\left\{\frac{s_m}{2} + \sum_{k=1}^{k=m}(-1)^k s_k\right\}; \\ (30)\qquad a_r &= \frac{2}{m}\left\{\frac{s_m}{2} - \sum_{k=1}^{k=m} s_k\cos\frac{r\pi}{m}k\right\}, r = 2, 4, 6, \cdots, m-1; \\ a_r &= -\frac{2}{m}\left\{\frac{s_m}{2} + \sum_{k=1}^{k=m} s_k\cos\frac{r\pi}{m}k\right\}, r = 1, 3, 5, \cdots, m-2. \end{aligned}\right\}$$

If in (4) a_m be replaced by $\frac{a_m}{2}$, the three forms (30) for the coefficients a_r can be expressed in the single form

$$(31)\qquad a_r = (-1)^r\frac{s_m}{m} - \frac{2}{m}\sum_{k=1}^{k=m} s_k\cos\frac{r\pi}{m}k.$$

Suppose now m is *even*. Then we obtain in manner parallel to the case m odd:

$$(28')\qquad M = -\frac{1}{m}\sum_{t=1}^{t=\frac{m}{2}}\sum_{k=1}^{k=m} s_k\cos\frac{(2t-1)\pi}{m}k$$

$$(29')\qquad = \frac{s_m}{2}.$$

*In practical computation, the two values found for M serve as an invaluable check.

Digitized by Google

The coefficients a_n then become

$$(30') \quad \left.\begin{aligned} a_m &= \frac{1}{m}\left\{\frac{s_m}{2} - \sum_{k=1}^{k=m}(-1)^k s_k\right\}; \\ a_r &= \frac{2}{m}\left\{\frac{s_m}{2} - \sum_{k=1}^{k=m} s_k \cos\frac{r\pi}{m}k\right\}, \; r = 2, 4, 6, \cdots, m-2; \\ a_r &= -\frac{2}{m}\left\{\frac{s_m}{2} + \sum_{k=1}^{k=m} s_k \cos\frac{r\pi}{m}k\right\}, \; r = 1, 3, 5, \cdots, m-1. \end{aligned}\right\}$$

From (30′) it is clear that (31) gives the coefficients a_n also when m is even. Having expressed the coefficients a_n in terms of the definitely known quantities s_k, the function (4) is then completely determined.*

In practical computation, the labor involved consists primarily in computing the values of $\sum_{k=1}^{k=m} s_k \cos\frac{r\pi}{m}k$ for all values of $r \leqq m$. This requires the m^2 values of $s_k \cos\frac{r\pi}{m}k$. But it is readily observed that all the m^2 values of $\cos\frac{r\pi}{m}k$ are given by only $\frac{m-1}{2}$ of them (if m is odd). Furthermore, s_k is a "constant multiplier" for each of the values of $\cos\frac{r\pi}{m}k$, $r \leqq m$. In practice, further use can be made of the simple relations between the trigonometric functions to materially shorten the work. The usual case arising for interpolation, is to obtain the frequencies for midway divisions of the sub-intervals. The computation made necessary for this is about equivalent to that in obtaining the coefficients a_n from the given frequencies a_k.

* If the number of sub-intervals of $[a, b]$, m, be indefinitely increased, it may be shown without difficulty, as is to be expected, that the function (4) becomes a Fourier series.

Digitized by Google

REVIEWS AND NOTES.

Poverty and Social Progress. By Maurice Parmelee, Ph.D. New York: The Macmillan Company, 1916.

Professor Parmelee has undertaken in this book of nearly five hundred pages to "give a comprehensive survey of the problems of poverty which shows the one-sided character of many of the explanations of its causation and which will furnish the starting point of an effective program of prevention." This promise of the preface has been fulfilled in a more satisfying manner than in any American work on this subject. In fact, there is none to be compared with it in its broad comprehensiveness. Like most books of moderate size that achieve a comprehensiveness, it cannot be intensive within its physical limits. It must be, and is, summary in its treatment.

After a short introduction, the study falls into two natural subdivisions: the causes and conditions of poverty; and the remedial and preventive measures, both those in use and those suggested.

Causes and conditions are summarized as follows: Factors biological, such as inheritance, the pathology of body and of mind; factors involved in the distribution of wealth and income; factors inherent in the present organization of labor, *e. g.*, unemployment, sweating, child-labor; the relationship of population, the means of subsistence; and political and domestic maladjustment.

In his survey of remedies and prevention, Professor Parmelee discusses humanitarianism, philanthropy, and eugenics, subjecting the claims of each as a remedy of the evil to a sharp criticism. Thrift, social insurance, and pensions are shown to be limited in their effectiveness, both as remedies and preventives. True prevention must, of course, be the aim of society in its treatment of poverty. Social legislation tends in that direction, though with only partial success. A study of the economic organization of modern society reveals certain basic facts. These are: the enormous concentration of wealth and income of a few; the instability of industry; the positive waste of economic goods and the negative waste caused by inefficient employment of human energies. Only fundamental, deep-striking reforms of these defects of our economic system can be relied upon to help us out of this perpetual slough of despond. Social machinery which will raise wages, regulate the labor supply, redistribute income from ownership of property, eliminate waste due to instability of industry and to lack of full utilization of the ability of the worker, will be effective in a large degree. The alternatives of highly regulated competition, and of collectivist industrial democracy afford two possible basic principles upon which to elaborate such machinery. Such machinery would require a political democracy more efficient than any we now know; and it would imply a development of the "world state" or some step in that direction.

Digitized by Google

As for the other half of the title of the book, "social progress," our author defines its aim as the coöperative attainment of the well-being of all mankind; its standard of value is frankly and nobly hedonistic. Human well-being is the coming of the "normal life" for all, or the vast majority, of mankind; the spontaneous expression of human nature. Through democracy alone will come the abolition of poverty, the greatest step toward the state of the "normal life."

The statistician is of course interested in the use which Professor Parmelee makes of the statistical material upon which in part his argument rests. This material is found in the chapters on the distribution of wealth and of income; on the extent of poverty; on unemployment; on the sweating system; on child and woman labor; strikes; on the growth of population and the increase of wealth; on domestic and matrimonial maladjustment.

A careful survey of the use Professor Parmelee has made of his statistical data gives the reader the impression that he has in no case (except one) drawn a more than conservative conclusion from the voluminous and highly heterogeneous material. He does not usually make a severe criticism of his sources; but he refuses to admit that Hunter's facts prove the contention the latter makes as to the extent of poverty. He regards this conclusion of Hunter's, however, as borne out in large part by corroboratory data. Professor Parmelee's careful use in every case of the data offered by the specialists relieves him of the need of so rigorously critical a treatment. In other words, he accepts the data as somewhat inexact, but does not attempt to do what too many students (notably Hunter) have done, namely to make exact inductions from inexact statements of the facts.

One point which will interest the student of the problem of immigration is the conclusion that Professor Parmelee comes to regarding the relationship between immigration and the birth rate of the original English-speaking stock of the United States. On the basis of the estimate in the census publication, "A Century of Population Growth," he admits the probability of Walker's well-known statement in this matter, and adds that immigration has not increased population in the United States but may even have checked its growth. That is a conclusion that might be difficult to justify from logical consideration of his data, as well as upon other grounds.

In conclusion, one point arises for question rather than criticism. What does Professor Parmelee really mean by the coming of the "normal life" as the "object of social progress?" To define the "normal life" as the "spontaneous expression of human nature" is good philosophy after the style of Plato. It lacks the objectivity that we are demanding today. The definition seems to the writer less definite than the term defined. The lower limit of the social conditions of the normal life are by implication discussed in the book. We could have desired our author formally to set at least a theoretical norm that would give us more than a hint of the concrete meaning both of "spontaneous development" and of "human nature." But has this more than an academic importance?

C. E. Gehlke.

Western Reserve University.

Digitized by Google

Tables: To Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations. Truman Lee Kelley, Austin, Texas. Bulletin of the University of Texas.

In these tables Professor Kelley has supplied a substitute for most of the numerous arithmetic operations which occur in recurring fashion in the computation of partial coefficients of correlation. In computing the coefficients of regression, the indices of higher order are obtained from those of the first order by means of three fundamental equations of the following form:

$$r_{12.34\cdots n} = \frac{r_{12.34\cdots(n-1)} - r_{1n.34\cdots(n-1)} \cdot r_{2n.34\cdots(n-1)}}{\sqrt{1-r^2_{1n.34\cdots(n-1)}}\sqrt{1-r^2_{2n.34\cdots(n-1)}}}, \tag{1}$$

$$\sigma_{1.23\cdots n} = \sigma_1 \sqrt{1-r^2_{12}}\sqrt{1-r^2_{13.2}}\cdots\sqrt{1-r^2_{1n.23\cdots(n-1)}}, \tag{2}$$

$$b_{12.34\cdots n} = r_{12.34\cdots n}\frac{\sigma_{1.34\cdots n}}{\sigma_{2.34\cdots n}}. \tag{3}$$

It is to be noted that the formula (1) for obtaining a partial coefficient of correlation of any order from those of the next lower order is of the same form for all orders. To this end Professor Kelley has computed the values of $\frac{1}{\sqrt{1-a^2}\sqrt{1-b^2}}$ and $\frac{ab}{\sqrt{1-a^2}\sqrt{1-b^2}}$ for all values of a and b from 0 to 1 for equal differences of .01—in all, 20,000 tabular entries. Using equation (1), these tables facilitate the computation of successive orders of partial coefficients, starting with total coefficients of correlation.

The author gives another table for values of r, r^2, $1-r^2$, $\sqrt{1-r^2}$, $\frac{1}{\sqrt{1-r^2}}$, $\log\sqrt{1-r^2}$, $\log\frac{1}{\sqrt{1-r^2}}$, for all values of r from 0 to 1 for equal differences of .01. Using equation (2), this facilitates obtaining the standard deviation of any order from the standard deviation of lowest order. From these the regression coefficients are obtained at once by formula (3). The procedure involves getting first r, then σ, then b.

In his preface, the author promises the extension of the tables for smaller subdivisions of the variables, carried out also to further places of decimals. Though such tables would be more serviceable at the extreme values of the variables, it is questionable whether or not they would lead to more substantial results in forming conclusions based upon the coefficients of correlation. A difference of .001, for example, in any coefficient of correlation, is utterly meaningless to anyone.

In his discussion on "The Function of Partial Coefficients or Correlation and Regression Equations," Professor Kelley illustrates the use of his tables by several examples in detail. He presents a very convenient tabular form for the practical computation of regression equations, with a

Digitized by Google

detailed discussion of approximations to the regression equation for cases involving a large number of variables.

The presentation, however, adopts poor pedagogic procedure. The apparent implicit confidence in the value of these partial coefficients of correlation, though quite orthodox, seems calculated to create in the mind of the student of statistics a blind faith in the subtle power of this form of analysis. This would tend to make the use of coefficients of correlation simply mechanical, and to that extent eliminate the supreme advantage of naive statistical thought. One would fail to realize that the regression equation is not the final word in any statistical analysis. At best, it is but a convenient approximate representation of the relation between a variable and other variables which partially determine it. Though the tables of themselves are valuable, the author's comments would tend to be misleading to one who is unfamiliar with the precise arbitrary nature of a regression equation. Moreover, regressions are sometimes used for purposes of interpolation for such values of the variables where the actual regression equation is anything but linear. This should be guarded against. Such comments would of course lose their force when applied to what we may term regression hyper-surfaces, and the associated correlation ratios.

The author vaguely refers to $\sqrt{1-r^2}$ as the "coefficient of independence," and to $\frac{1}{\sqrt{1-r^2}}$ as the "measure of importance," "measure of significance," etc., of variables, though r is the correlation with but one other variable. In the second part of a regression equation a variable may be said to be "important" *by definition* in proportion to the size of the weights. It is poor pedagogy to use verbal interpretations of loose significance for functions of precise mathematical form. The latter is at best but an arbitrary means for the classification of the data.

Professor Kelley states the theorem that the determination of the regression equation on the principle of least squares is equivalent to that of determining a linear function of the independent variables which will yield values that correlate most highly with the corresponding values of the dependent variable itself. He refers the student to Yule's "Introduction" for an elementary proof. On referring to that book, however, the student will find nothing in the nature of a proof, though he may be satisfied by a casual statement of the author, without proof, to that effect.

The last sentence of the author's remarks is somewhat cryptic. One would be curious to discover the intended connotation.

HARRY LANGMAN.

New York City.

Digitized by Google

STATE CENSUSES IN 1915.

Ten states took a census of the population in 1915. The published returns for these censuses are not entirely satisfactory from a statistical standpoint, however, as pointed out in Bulletin 133, Bureau of the Census, 1916. Methods of enumeration, tabulation and publication, and the dates upon which the enumerations were made, differ widely.

In New York State, the published count of the population is believed by competent observers to be far short of the actual number. In New York City several hundred thousand persons were excluded from the census whose births, marriages, deaths, and sicknesses are reportable. The group of excluded persons were: (1) persons in federal reservations, and in navy yards, army posts, marine hospital stations; (2) guests of hotels; (3) inmates of institutions; (4) residents of the city temporarily absent; (5) persons, like day laborers, who regularly leave the city for out-of-town employment during the summer months. At a slight additional cost in money and pains, the New York State Census, taken primarily for legislative apportionment purposes, might have been made to serve also the health and social agencies of the state.

In New Jersey, no specific indictments of the state census were made. The general dissatisfaction with the meagre results of the enumeration was expressed, however, in the repeal of the enabling legislation by Senate Bill No. 130, chapter 34, Laws of 1916.

The bill under which the census was taken was enacted April 12, 1905, for the enumeration of that year. The New Jersey 1915 Census cost very nearly $100,000; $85,330 was paid to enumerators and supervisors. A voluntary committee representing the social and philanthropic agencies of the state had tried to have the scope of the enumeration and tabulation program of the census amended so as to provide data useful in the study of population movements which had taken place since 1910. No action by the state authorities followed this committee's recommendation, however.

Francis A. Walker in his article on the Eleventh Census of the United States (*Quarterly Journal of Economics*, Vol. 2, 1888, pp. 135–161) suggested a Federal Quinquennial Census. He indicated the failure of the provision of the March 3, 1879, census law for aiding states to take censuses at dates intermediate between the United States enumerations. Section 22 of this law stipulated that if state enumerations were taken on a date mean between the dates of the federal censuses, on schedules and forms conforming in all respects to those of the federal census, the Secretary of the Treasury was empowered to pay to the governors of such states a sum equal to 50 per cent. of the amount paid enumerators and supervisors in that state at the time of the next preceding federal census. A subsidy of this kind for the New Jersey 1915 Census would have more than paid for the costs of the extra tabulations asked for by the social workers of the state. It is sincerely hoped that provision for state censuses, and for federal coöperation in such enumerations, will be made in the Fourteenth Census

Digitized by Google

Bill. The following brief tabulation from Bulletin 133, Bureau of the Census, summarizes the state census situation in 1915:

STATES TAKING AN INTERDECENNIAL CENSUS UNDER STATE CONTROL IN 1915.

State.	Date of State Census.		Official in Charge of Enumeration.
	First.	Last.	
New York	1698	June 1, 1915	Secretary of State.
Rhode Island	1708	Mar. 1, 1915	Supt., Bureau of the Census.
New Jersey	1726	June 1, 1915	Secretary of State.
Massachusetts	1765	Apr. 1, 1915	Director, Bureau of Statistics.
Iowa	1838	Dec. 31, 1914	Sec'y, Executive Council.
Kansas	1855	Mar. 1, 1915	Sec'y, State Board of Agriculture
Florida	1895	July 1, 1915	Commissioner of Agriculture.
South Dakota	1895	May 1, 1915	Superintendent of Census.
North Dakota	1905	Apr. 1, 1915	Secretary of State.
Wyoming	1905	Apr. 5, 1915	Secretary of State.

E. W. KOPF.

New York City.

"*Interest Tables for Small Loans*," by Arthur H. Ham. The Spectator Company, New York. 52 pp. Price $4.00.

This small volume from the pen of Mr. Ham, Director of the Division of Remedial Loans of the Russell Sage Foundation, contains a vast amount of valuable matter in a small space. There are two sets of tables in the volume, the first showing the amount of interest at the rates of 1, 1½, 2, 2½, 3 and 3½ per cent. per month on sums of 50 cents to $300 for periods of one day to thirty days. In the second place there are tables showing the amount of interest at these rates on loans of $10 to $300, payable in 4, 6, 8, 10 and 12 equal monthly installments. In addition to the tables, there is an appendix containing the formulae for determining annual interest rate, amount of the interest charged, annual discount rate, and amount of the discount charge on loans payable in equal periodic installments computed on unpaid balances and not compounded. This volume should be of considerable service to lenders on chattel mortgages and wage assignments, to pawn-brokers, and to small money lending agencies. The book is attractively prepared and is arranged so as to make easy a consultation of the tables.

WM. B. B.

Digitized by Google

COMMITTEES.

The following officers and committees have been appointed since the Proceedings of the Annual Meeting were published:

Assistant Secretaries:

Robert E. Chaddock, New York, N. Y., and Henry J. Harris, Washington, D. C.

Associate Editors:

Louis I. Dublin, Julius H. Parmelee, Carl J. West, and John Cummings.

Committee on Fellows:

Wesley C. Mitchell, chairman, term expires 1918.
Adna F. Weber, term expires 1921.
Frederick L. Hoffman, term expires 1920.
E. Dana Durand, term expires 1919.
Charles H. Verrill, term expires 1917.

Committee on Federal Statistics:

John Koren, chairman, Thomas S. Adams, Leonard W. Hatch, Frederick L. Hoffman, and N. I. Stone.

Committee on State Statistics:

E. Dana Durand, chairman, Robert E. Chaddock, Charles F. Gettemy, Henry J. Harris, and E. W. Kemmerer.

Committee on Municipal Statistics:

LeGrand Powers, chairman, Frederick A. Cleveland, Fred C. Croxton, Louis I. Dublin, and Edward M. Hartwell.

Committee on Business Statistics:

Melvin T. Copeland, chairman, Julius H. Parmelee, H. S. Person, Warren M. Persons, and I. M. Rubinow.

Committee on International Statistics:

Royal Meeker, chairman, Roger W. Babson, Miles M. Dawson, Irving Fisher, and Wesley C. Mitchell.

Committee on Finance:

Harvey S. Chase, chairman, Walter E. Weyl, and S. B. Pearmain.

Committee on Library:

Roger W. Babson, chairman, Edmund E. Day, and Henry J. Harris.

Nominating Committee:

Thomas S. Adams, chairman, William Z. Ripley, and George P. Watkins.

Auditing Committee:

Martin J. Shugrue and Roswell F. Phelps.

CARROLL W. DOTEN,
Secretary.

Digitized by Google

Digitized by Google

AMERICAN
STATISTICAL ASSOCIATION

NEW SERIES, No. 119. SEPTEMBER, 1917.

ON THE ALLEGED INCREASE OF CANCER.

By Walter F. Willcox, Ph.D., LL.D., *Professor of Economics and Statistics, Cornell University.*

The statistical study of cancer as a cause of death, which is the subject of this paper, should not be confused with that of cancer as a disease. Whether the disease is increasing can be answered only by trustworthy morbidity statistics. As no country or city in the world has such statistics, no significant statistical study of cancer as a disease has been or can be written. Hence the present paper is concerned only with the alleged increase of cancer as a cause of death. But at the start it should be conceded that, strong as the argument is for the increase of cancer mortality, that for the increase of cancer morbidity is probably stronger, *i. e.*, the case fatality is probably less than formerly. This would follow from the increasing proportion of operations to cases, in some of which the disease never recurs, in others of which its progress is stayed, affording a longer interval for the occurrence of a death from some other cause. The same conclusion is supported by other improvements in methods of treating cancer, all tending to prolong the patient's life after the onset of the disease and to increase the chance of an intervening death from some other cause.

This paper, however, is concerned only with the question, Is cancer mortality increasing? In nearly all parts of the civilized world the reported mortality is increasing rapidly and this is the basis for the prevalent opinion that the real mortality is increasing. General statements about improvements in the diagnosis of cancer and the increasing proportion of

Digitized by Google

elderly persons especially liable to the disease warrant the opinion that the true increase is less than that shown by the figures, an opinion held by all students of the subject, even those who seem at times to accept the figures at their face value. But these general statements alone do not justify the denial of any increase and leave the problem as one inviting further investigation.

Studies in this field are of two main types, which may be distinguished as intensive and extensive. In the case of an intensive study the figures examined speak for only a limited area, usually a city, and are analyzed in various ways, in order to disentangle their true meaning and to confirm or correct one inference by another. In the case of an extensive study, the figures speak for many different districts, and reliance is placed upon the corroboration derived from similar indications in many or all.

Nearly all studies of the increase of cancer have been made by physicians who have done little work in statistics, but two careful examinations have been made by statists of high repute and are so much more thorough and significant than any of the others that they stand in a class by themselves. One, written by Messrs. King and Newsholme and published in 1893, is a model of an intensive investigation, mainly into evidence derived from the city of Frankfort in Germany.* The other, written by Mr. F. L. Hoffman and published in 1915, is the only example with which I am acquainted of an extensive investigation.† But two other international compilations of vital statistics, including statistics of cancer, have recently been published‡ and a large body of American material for the decade 1900–1909 and the year 1914 has been brought together.§ All these contain uninterpreted figures which can

* "On the Alleged Increase of Cancer," in Royal Society, *Proceedings*, liv, pp. 209–242. Read May 4, 1893. To be cited henceforth in this article as *King and Newsholme*. Republished in Institute of Actuaries, *Journal*, XXXVI, part II (July, 1901), pp. 120–150.

† *The Mortality from Cancer Throughout the World*, The Prudential Press, Newark, N. J., 1915. To be cited henceforth in this article as *Hoffman*.

‡ *Statistique Internationale du Mouvement de la Population*, in two volumes, Paris, Imprimerie Nationale, 1907 and 1913. To be cited henceforth in this article as *March*. *Statistique Démographique des Grandes Villes du Monde, 1880–1909*, in two parts. (Communications statistiques, Bureau municipal de Statistique d'Amsterdam, nos. 33 and 40.) Amsterdam, Johannes Müller, 1911 and 1912. To be cited henceforth in this article as *Falkenburg*.

§ United States Bureau of the Census, *Mortality Statistics, 1909*, Washington, 1912; and *Mortality from Cancer and Other Malignant Tumors in the Registration Area of the United States, 1914*, Washington, 1916.

Digitized by Google

be made to throw no little light upon the alleged increase of cancer mortality.

In view of this situation, the objects at which I aim are:

1. To review the intensive study of Messrs. King and Newsholme and to present new evidence from Frankfort for the period 1890–1913 by which their argument may be tested.

2. To summarize the results of other intensive studies which have been published since 1893.

3. To make a new extensive study by reviewing and interpreting the large body of evidence recently published in the international and American sources with no fuller interpretation than that found in Mr. Hoffman's volume.

Such a paper seems the more opportune for two reasons: first, because the two most significant studies reach diametrically opposite conclusions, King and Newsholme alleging that "the increase in cancer is only apparent and not real,"* and Hoffman maintaining "that the mortality from cancer is increasing at a more or less alarming rate throughout the entire civilized world"†; and secondly, because little, if any, attention has hitherto been given to the light thrown on this problem by recent international compilations.

The competent statistical study of this question began in 1893 with the paper by Messrs. King and Newsholme. Mr. George King, one author, was Secretary of the Institute of Actuaries and Dr. Arthur Newsholme, the other, was the author of the standard English treatise on vital statistics. This article, written in coöperation by men both adequately equipped in statistics and one equipped in medicine, is probably the most significant statistical discussion of the problem which appeared before 1915 and therefore I shall begin with an examination of its arguments in the light of the additional evidence made available since it was written.

To prove that there is no increase in cancer mortality, all of the apparent increase must be explained by an appeal to other causes. Messrs. King and Newsholme were the first students of the problem who trod this difficult road and they have not found imitators. It is in the choice of a method rather than in

* *King and Newsholme*, p. 228.
† *Hoffman*, p. 218.

Digitized by Google

the conclusiveness of the results that the main value of their paper lies.

Their main conclusion and the one with which we are primarily concerned was that "the increase in cancer is only apparent and not real." Their argument shows that this statement refers only to cancer mortality and is in no wise inconsistent with an increase of cancer morbidity offset by a diminishing case fatality.

Their subsidiary conclusions supporting this main conclusion were two:

1. A part of the increase in the crude cancer mortality is due to the increasing average age of the population and is eliminated when standardized death rates are computed.

2. In Frankfort-on-the-Main, the only population group in which the annual budget of deaths from cancer at each age and for each sex had long been carefully classified by the organs or parts of the body primarily affected, "the increase in cancer is only apparent and not real and is due to improvement in diagnosis and more careful certification of the causes of death. This is shown by the fact that the whole of the increase has taken place in inaccessible cancer difficult of diagnosis, while accessible cancer easily diagnosed has remained practically stationary."*

This argument was apparently regarded by the authors as decisive and upon it they rested their case.

The second of these two main arguments was supported by two minor arguments, which may first be reviewed in the light of the new evidence.

1. The increase of cancer mortality between 1860 and 1891 among men insured in the Scottish Widows' Fund Life Assurance Society, the only group of insured persons for which the requisite information could be obtained, was much less rapid than its increase in the general population of Scotland. This difference was explained by supposing that, when the figures began, the diagnosis of cancer among insured persons was more accurate and among them, in consequence, the effect of improvement of diagnosis during the thirty years was less than in the general population.

* *King and Newsholme*, p. 228.

Digitized by Google

2. Cancer mortality was lower but increased faster among men than among women. This was ascribed to the greater proportion of cases of inaccessible cancer among men, the greater obstacles to accurate diagnosis in such cases, and the more rapid improvement in the diagnosis of inaccessible cancer.

Their argument from insurance experience is illustrated by the following figures compiled from those in their article.

DEATH RATE FROM CANCER PER 100,000 MALES AT LEAST TWENTY-FIVE YEARS OF AGE, DISTRIBUTED IN AGE GROUPS AS IN ENGLISH LIFE TABLE NO. 3.

Period.	Insured in Scottish Widows' Fund Life Assurance Society.	In Total Population of Scotland.
1860–1866	60	70
1867–1873	83	82
1874–1880	75	101
1881–1887	85	115

The increase of the death rate among the general male population was 45 per 100,000, or 64 per cent., that among the insured males was 25 per 100,000, or 42 per cent. Mr. Hoffman quotes some additional evidence furnished by the medical officer of the same Society,* the figures of which, it will be noted, do not agree entirely with those published by Messrs. King and Newsholme. I learn from the actuary of one of the largest American insurance companies that probably no evidence on this point has been published by any American company. But even if it were buttressed by much additional material, I cannot attach great weight to this argument, because of the uncertain influence of selection in any class of insured persons. During the thirty years covered by the experience of the Scottish insurance company, for example, it may have extended its business among classes somewhat less subject to cancer than its earlier policy holders or the protective influence of the company's medical selection may have diminished.

The second argument advanced by Messrs. King and Newsholme is based upon the lower rate and more rapid increase in the cancer mortality of men. Among males cancer usually

* *Hoffman*, pp. 84, ff.

Digitized by Google

attacks an inaccessible organ or part of the body and for that reason is difficult to diagnose correctly even at the time of death, a difficulty which has decreased with the improvements of modern medicine and surgery. Thus, accepting as correct their classification of deaths from cancer in Frankfort into the two groups of accessible and inaccessible, 97 per cent. of the male deaths and only 63 per cent. of the female deaths ascribed to such cancers occurred where the seat of the disease was inaccessible.* Their argument is illustrated by the following figures computed from those in their paper.

DEATH RATE FROM CANCER PER 100,000 PERSONS OF SEX INDICATED AND AT LEAST TWENTY-FIVE YEARS OF AGE DISTRIBUTED IN AGE GROUPS AS IN ENGLISH LIFE TABLE NO. 3.

Country.	1860.	1890.	Increase.	
			Amount.	Per Cent.
Males.				
England	59	145	86	146
Scotland	67	174	107	160
Ireland	59	92	33	56
Females.				
England	119	210	91	77
Scotland	116	220	104	90
Ireland	72	111	39	54

The foregoing figures show that in all these divisions the female rates at both dates were much higher, that in England and Scotland the actual increase for the two sexes is about the same and the relative increase for men much greater, but that in Ireland the same facts do not appear. What do the recently published compilations show about the comparative increase of cancer in the two sexes? *Hoffman* alone gives death rates by sex. These he gives for each of the New England states, for Massachusetts since 1856, for Rhode Island since 1871, for Connecticut since 1879, for New Hampshire since 1887, for Maine since 1892, and for Vermont,

* The data involved will be found in the Appendix, Table VIII.

Digitized by Google

1871–1896. In 2 of these states the absolute increase and in 4 the relative increase of cancer mortality was greater among men than among women. A summary of his figures for the first 5 states since 1896 by sex yields the following results:*

Period.	Death Rate per 100,000 Population.		Increase.				Ratio of Male Rate to Female (=100).
			Amount.		Rate.		
	Male.	Female.	Male.	Female.	Male.	Female.	
1896–1900	46.2	90.3					51.2
1901–1905	53.6	100.9	7.4	10.6	16.0	11.7	53.1
1906–1910	61.5	110.9	7.9	10.0	14.7	9.9	55.5

These figures show that the rate of increase in cancer mortality among men has been somewhat more rapid than among women and that during the period the ratio of male mortality to female increased 4.3 per cent.

For the Federal registration area as constituted in 1900† the cancer death rates by sex were:

Year.	Death Rate per 100,000 Population.		Increase.				Ratio of Male Rate to Female (=100).
			Amount.		Rate.		
	Male.	Female.	Male.	Female.	Male.	Female.	
1900	47.0	80.7					58.2
1915	72.3	111.9	25.3	31.2	53.8	38.7	64.6

indicating that among men the increase was less, but the rate of increase greater, than among women. The ratio of male mortality to female increased 6.4 per cent. in fifteen years.

Among the 26 American cities for which Mr. Hoffman gives death rates by sex there are 10 in which the absolute increase and 19 in which the relative increase of the cancer mortality

* For the figures on which these rates are based see Appendix, Table I.

† Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, District of Columbia, Michigan, and Indiana. See Appendix, Table II. For our problem the figures of "the registration area" are without value, because of that area's rapid expansion since 1900.

Digitized by Google

is greater among men. The figures by sex since 1891 are given for 14 American cities.* They yield the following summary:

Period.	Death Rate per 100,000 Population.		Increase.				Ratio of Male Rate to Female (=100).
			Amount.		Rate.		
	Male.	Female.	Male.	Female.	Male.	Female.	
1891–1895	39.4	71.9					54.9
1896–1900	45.8	77.7	6.3	5.8	15.9	8.1	59.0
1901–1905	52.1	85.8	6.4	8.1	14.0	10.4	60.8
1906–1910	61.8	95.5	9.7	9.7	18.7	11.3	64.7

Thus, the cancer mortality of men in American cities has increased faster than that of women and the change has gone on more rapidly than in the American states. In fifteen years the ratio of male mortality to female has increased 9.8 per cent.

Among the 30 foreign countries for which Mr. Hoffman gives death rates by sex, there are 12 in which the absolute increase and 20 in which the relative increase of cancer mortality is greater among men than women and in the minority of 10 are 5 small Australasian divisions. He gives the figures by sex since 1895 for England and Wales, Ireland, Norway, Bavaria, Italy, New South Wales, Victoria, South Australia, Queensland, Tasmania, New Zealand, Bermuda, and Jamaica. By combining the results the following summary has been obtained.†

Period.	Death Rate per 100,000 Population.		Increase.				Ratio of Male Rate to Female (=100).
			Amount.		Rate.		
	Male.	Female.	Male.	Female.	Male.	Female.	
1896–1900	57.2	76.9					74.4
1901–1905	63.8	82.3	6.6	5.4	11.5	7.0	77.5
1906–1910	72.2	89.4	8.4	7.1	13.2	8.6	80.8

* Boston, Springfield (Mass.), Providence, Hartford, New Haven, New York, Philadelphia, Washington, Richmond, Augusta, Cleveland, Cincinnati, St. Louis, and Nashville. Appendix, Table III.

† Appendix, Table IV.

Digitized by Google

In this case the tendency of the male rate to overtake the female is more clearly marked, for the amount as well as the rate of increase is greater among the men. The ratio of male mortality to female increased 6.4 per cent. As each rate in the table is based upon more than 100,000 deaths from cancer, the steady decrease in the difference between the sexes is of more importance for the argument than that shown in the preceding tables.

Among the 33 foreign cities for which Mr. Hoffman gives the mortality by sex there are 24 in which the absolute increase and 26 in which the relative increase of the cancer death rate is greater among men. The figures by sex since 1896, given for 13 foreign cities separately and the cities of Denmark collectively, yield the following summary.*

Period.	Death Rate per 100,000 Population.		Increase.				Ratio of Male Rate to Female (=100).
			Amount.		Rate.		
	Male.	Female.	Male.	Female.	Male.	Female.	
1896–1900	80.4	107.6					74.7
1901–1905	90.3	115.0	9.9	7.4	12.3	6.9	78.5
1906–1910	99.4	120.1	9.1	5.1	10.1	4.4	82.8

The ratio of male mortality to female has increased 8.1 per cent.

All the new evidence supports the conclusion of Messrs. King and Newsholme that the death rates from cancer in the two sexes are approaching equality as a result of the more rapid increase among men.

Having reviewed the new evidence in these two subsidiary arguments, we turn now to the second of the main arguments advanced by Messrs. King and Newsholme, an argument supported by this difference between the cancer mortality of the two sexes. This alleges that the increase in cancer mortality is confined to mortality from inaccessible cancer. Their evidence on this point is derived from the figures of Frankfort-on-the-Main. They seem to have accepted the Frank-

* Appendix, Table V.

Digitized by Google

fort returns without any attempt to investigate the accuracy of the classification according to the part of the body affected by a study on the ground both of the original material and of the methods of treating it, a precaution which might well have been taken when so much depended on its accuracy. The able statists of Frankfort probably did not attach to this classification as much weight as the English interpreters. Certainly they abandoned it in 1891 and the Frankfort director, in a letter to me on the subject, volunteers the statement that in the opinion of German statists the diagnosis of cancer is not entirely trustworthy and that physicians out of regard for the family sometimes certify a death really due to cancer to be due to some less distasteful cause, like tumor.

Notwithstanding these reasons for giving less weight than our authors do to the Frankfort figures, it seemed wise to bring them down to date, if possible,—the more so, because in the 24 years since their article was published little new evidence has appeared based on a classification of deaths from cancer as located in accessible or inaccessible parts of the body. It was found, as mentioned above, that the Frankfort statistical office had abandoned the classification after 1891; consequently it was necessary to make a special tabulation of the figures since that date.* In order to insure comparability between the later and earlier figures, the returns for 1888 and 1889, which Messrs. King and Newsholme had compiled, were recompiled and their treatment of a few doubtful groups followed.

Some writers have objected to the classification of deaths from cancer according as the seat of the disease was accessible

* The needed sum was given by Dr. James Douglas at the suggestion of Prof. James Ewing, at whose request my paper was begun and who has aided me throughout its progress. Without his encouragement I should hardly have ventured so far into the field of medical statistics. I am greatly indebted to Director Busch of the Frankfort statistical office, who arranged to have his subordinates compile the data, and to Prof. R. M. Woodbury, then a travelling Fellow of Cornell University pursuing studies in economics and statistics at the University of Berlin, who went twice from there to Frankfort to explain exactly what was wanted and to review the results. His later services in Ithaca were also of much help in this part of the problem. The manuscript tables obtained from Frankfort have been deposited with the Library of the Surgeon-General's Office in Washington, where they may be examined by any persons desiring to test my inferences from them. The figures for 1860 to 1889, inclusive, are in *King and Newsholme;* those for 1890 are in the *Stat. Mittheilungen* of Frankfort for that year. But some of the certificates for 1890 and 1896 are now lacking and hence the returns for those two years have been disregarded. In Appendix, Tables VI–X, will be found all of Messrs. King and Newsholme's Frankfort material, each table supplemented by additional figures so as to add five new periods between 1888 and 1913.

Digitized by Google

or inaccessible, alleging either that a third class of intermediate cases should have been recognized or that certain organs or parts considered by Messrs. King and Newsholme as accessible should have been put in the inaccessible class or *vice versa*. In reply to the first objection, it is clear that the introduction of a third class of intermediate cases would obscure the argument. For our purposes a dual classification with a few border line groups thrown to one side or the other according to where most of the cases in the group belong is far better than the alternative triple classification. Whether their classification could have been improved is a question of technical medicine into which fortunately there is no need to go. My present object being merely to bring their material down to date, I would have to follow their classification, even if I knew, as I do not, that it had serious defects. But, in passing, I may mention that their decisions have found remarkable confirmation in the statistics of the Census Bureau's special bulletin on *Mortality from Cancer in 1914*, mentioned on p. 702, footnote §. Thousands of American doctors were asked to state, in connection with each cancer death certificate which they had filled out, whether the diagnosis was based on clinical findings, or confirmed by surgical operation or autopsy. With the help of their answers, the cancer deaths were classified under the captions "Diagnosis reasonably certain," and "Diagnosis uncertain." To this point I shall recur later (pp. 749, ff).

The new tables for Frankfort extend the original ones so that they embrace the fifty-four years, 1860–1913, instead of only the thirty years, 1860–1889, and nine thousand deaths instead of three thousand.* The final result is shown in the following table:

* Compare Appendix, Tables VI–X.

Digitized by Google

FRANKFORT. ANNUAL DEATHS FROM CANCER IN 1,000,000 LIVING, AT LEAST TWENTY-FIVE YEARS OF AGE. POPULATION DISTRIBUTED IN AGE GROUPS ACCORDING TO ENGLISH LIFE TABLE NO. 8.—PERSONS.(a)

Period.	Accessible.	Inaccessible.	Position Undefined.	Total.
		MALES.		
1860–1866	126	1,118	359	1,603
1867–1873	88	1,421	137	1,646
1874–1880	14	1,913	363	2,290
1881–1887	35	1,865	305	2,205
1888–1893(b)	91	2,289	171	2,551
1894–1899(c)	66	2,459	104	2,659
1900–1904	51	2,742	118	2,911
1905–1909	68	2,791	119	2,981
1910–1913	119	2,522	130	2,771
		FEMALES.		
1860–1866	1,081	1,323	293	2,697
1867–1873	1,214	1,540	254	3,008
1874–1880	1,220	1,588	131	2,939
1881–1887	981	1,820	272	3,073
1888–1893(b)	1,275	1,959	174	3,408
1894–1899(c)	903	1,936	122	2,961
1900–1904	914	2,013	115	3,042
1905–1909	925	2,311	180	3,416
1910–1913	913	2,057	152	3,122

(a) Figures from 1860 to 1887, inclusive, derived from *King and Ntuchelme*, Table XVI. Figures for later years derived from manuscript tables from the Frankfort Statistical Office.
(b) Omitting the year 1890, for which the figures were not available.
(c) Omitting the year 1896, for which the figures were not available.

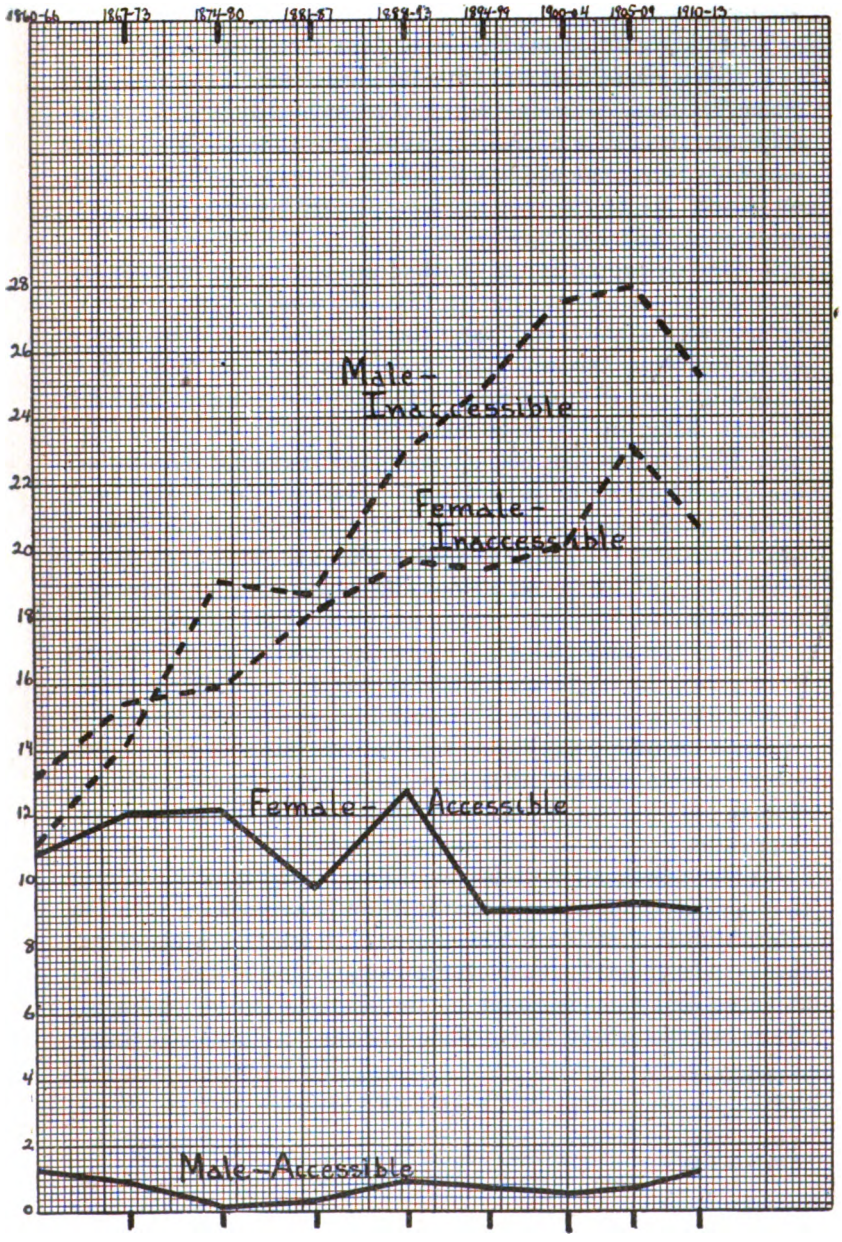
The general meaning of the table can be read more easily and clearly in the form of a diagram.* For the sake of simplicity the categories of "position undefined" for each sex and the total rates for each sex have been omitted and the diagram depicts, therefore, the conditions stated in the first two columns of the preceding table.

Both table and diagram show, each in its own way, that there has been no significant change in the small mortality from accessible cancer of men and there have been irregular changes followed by a remarkable uniformity in the mortality from accessible cancer of women. Almost all the increase occurred in the mortality from inaccessible cancer, although by a surprising exception that mortality in each sex decreased between the last two periods, while the mortality of men from

* See p. 713.

Digitized by Google

DIAGRAM I. DEATH RATES FROM CANCER, ACESSIBLE AND INACCESSIBLE, FRANKFORT-A-M.: 1860–1913.



Digitized by Google

accessible cancer increased. The argument of Messrs. King and Newsholme from the Frankfort figures is thus materially strengthened.

To this argument several objections have been made.

1. "It is quite doubtful whether all uterine or even vaginal cancer can be accurately diagnosed as such without an exploratory operation or a microscopical examination of the diseased parts."* The evidence in the Census Bureau's recent bulletin on cancer, to which reference has already been made (p. 711) leads to a different conclusion.

2. "A fair proportion of deaths from cancer are those of the skin, other parts of the mouth than the tongue, and other external parts of the body"† and these should have been classed as accessible. The objection overlooks the fact that these classes are not distinguished in the printed sources used by Messrs. King and Newsholme.

3. The authors "for reasons unknown" have not reëxamined the Frankfort data since 1893.‡ This overlooks the fact, already mentioned, that the method of publishing the Frankfort figures was radically changed at about the time the original article was published and serious difficulties put in the way of a reëxamination.

4. It is not clear whether correction was made for deaths from cancer among inmates of Frankfort hospitals who were non-residents of that city.§ Careful inquiry at the Frankfort registration office, supplemented by inspection of the manuscript returns, showed that they were not systematically excluded, but that the smallness of their number made them unimportant.

5. The classification of deaths in Frankfort "fails to conform to modern requirements in that it is not in accordance with the international classification of causes of death."‖ A classification is to be judged with reference to the object it was designed to serve and the character of the material to be classified. There is little reason to suppose that the international classi-

* *Hoffman*, p. 31.
† *Ibid.*
‡ *Ibid.*
§ *Ibid.*, footnote.
‖ *Hoffman*, p. 33.

Digitized by Google

fication is better adapted to the specific study of cancer death rates than another which may have had that object in view and certainly has been found valuable for it.

6. The results do not agree with those derived from the Frankfort data for 1906–1913 given in the official reports of the Frankfort Medical Society.* The latter are in conflict with those for the same years which I have derived from the official records of the Frankfort statistical bureau and already presented, and the differences appear irreconcilable. The figures of the Medical Society make no attempt at a classification into accessible and inaccessible and in consequence have little bearing upon the argument.

7. The standardized death rates from cancer in Frankfort, 1860–1889, by five periods of years show irregular changes, perhaps due to paucity of numbers, but they cannot be interpreted as affording no evidence of increase in the rates for accessible cases among women.† To this objection the supplementary figures already presented afford perhaps a sufficient reply.

These objections to the argument of Messrs. King and Newsholme do little to weaken its probative value. An objection of greater weight than these lies against their reliance upon standardized rates and will have to be considered later. In face of all the objections we can hardly accept their conclusion in the emphatic way in which it is stated. Neither the original argument nor my amplification of it, nearly doubling the period of observation and more than trebling the number of cases, can be held to justify completely their statement that "the increase in cancer is only apparent and not real and is due to improvement in diagnosis and more careful certification of the causes of death." The great merits of their paper lie in exemplifying the correct method of attacking the problem and showing beyond question that a measurable part of the increase in the mortality from cancer is due to a better diagnosis of the causes of death and another measurable part to the increasing proportion of persons who live on to ages when cancer is prevalent.

Let us now examine the first of their two main arguments,

* *Hoffman*, p. 31.

† Greenwood and Wood, "On Changes in the Recorded Mortality from Cancer and their Possible Interpretation," in *Proceedings of the Royal Society of Medicine*, VII, no. 1, pp. 119–70.

Digitized by Google

namely, that part of the increase is due to changes in the age and sex composition of the population and can be measured by the use of standardized death rates. No one will question that the increasing proportion of elderly persons in the population accounts for part of the increase in cancer mortality. But that this part can be measured by the help of standardized rates has been somewhat uncritically taken for granted in the United States and England. Various earlier writers had pointed out that cancer is preëminently a disease of adult and elderly life and that the proportion of the population at any moment living at ages when cancer is frequent is increasing with the lengthening of human life. But it was reserved for Messrs. King and Newsholme to put the argument into the necessary quantitative form. In computing standard death rates, they chose for the population a standard million based on English Life Table No. 3, Persons. To this they applied the cancer death rates for each sex and age group at each of two periods and compared the results. In such a computation the increase or decrease in the computed deaths is supposed to show by how much the cancer mortality would have increased or decreased, if there had been no change in the sex and age composition of the population but such changes as did occur in death rates.

This method of using standardized rates has long been followed in England, has recently been adopted in the United States, and is regarded by Messrs. King and Newsholme as a prime requisite to sound work in dealing with the present problem. They say: "In order to be useful, the materials . . . must be presented in such form that, at any rate for several intervals of years, the deaths from cancer and the population may be grouped according to sex and age"*; and it was because these conditions were satisfied nowhere but in the Frankfort statistics and those of the Scottish Widows' Fund Life Assurance Society that our authors relied so heavily upon that material. Their principle is a veritable bed of Procrustes. Its acceptance excluded at one stroke probably all other cancer statistics than those mentioned, and, if it is sound, the significant material is still so meager as to make an

*P. 212.

Digitized by Google

extensive study even now quite out of the question. Accordingly the principle should not be accepted without a more careful scrutiny than it has hitherto received. The method assumes that a standard population is needed but that it is a matter of indifference what standard is adopted. Thus Messrs. King and Newsholme say: "It is not of importance what standard is selected, because, all the observations being treated alike, the comparisons instituted between them will be entirely trustworthy."*

It is desirable to put this assertion to the test. The standards most likely to be in the mind of Englishmen writing in 1893 were the English Life Table No. 3 and the population of England in 1881, the date of the latest available census. A comparison of these two standards is made in Table XI of the Appendix and is expressed to the eye through the diagram on page 719.

The diagram shows clearly what the figures confirm and make precise, that the proportions actually living at the various ages in England in 1881 were very different from those shown by a life table. The life table assumes a population exempt from the disturbing factors of emigration and immigration and also from the distortion engendered by a rapidly increasing population, which swells inordinately the contingents belonging to the early years of life. Cancer is a cause of death almost unknown in infancy and childhood but common in adult and elderly years. Hence the standardized death rate from cancer obtained by using a standard in which adults and elderly persons are numerous must be much above the rate obtained by using a standard in which they are relatively few. For example, the standardized cancer mortality of Frankfort, 1860–1866, was 107 per 100,000, if the standard used is the English Life Table population, but only 70, or about two thirds as high, if the standard is the population of England and Wales in 1881.† To this objection it might be replied that the information needed is not the standardized rate in any single period but the increase between two periods and that the increase, absolute or relative, might be the same even though

* Pp. 211–12.
† See Appendix, Table XII.

Digitized by Google

the rates compared were
is unsatisfactory for two r
were the same, the relativ
be; and, secondly, as the
well as its amount varies
the agreement of two in
divergent standards are
hand, if the life table pop
in the standardized deatl
100,000 in 1860–1866 to 1
60 per 100,000, or 57 per c
in 1881 be the standard,
to 107 in 1888–1889, whic

From this argument it
of standardized death rat
inapplicable to the presen
too far. In most civiliz
population is slowly incre
death rate, especially in e
the birth rate. For exam
age of the population incr
1910,† or about four fifths
though less rapid, chang
other countries. Wherev
part of the increase in c
theoretically possible tha
Under these circumstanc
due to that cause must b
on the threshold. An esc
along the following path i

To the population of a
of a period the correspon
of the period are applie
being the number which w
tion of the population re
a "corrected" death rate
due only to the changes i

* Appendix, Table XIII.
† United States Bureau of the Census, *Ti*

Digitized by Google

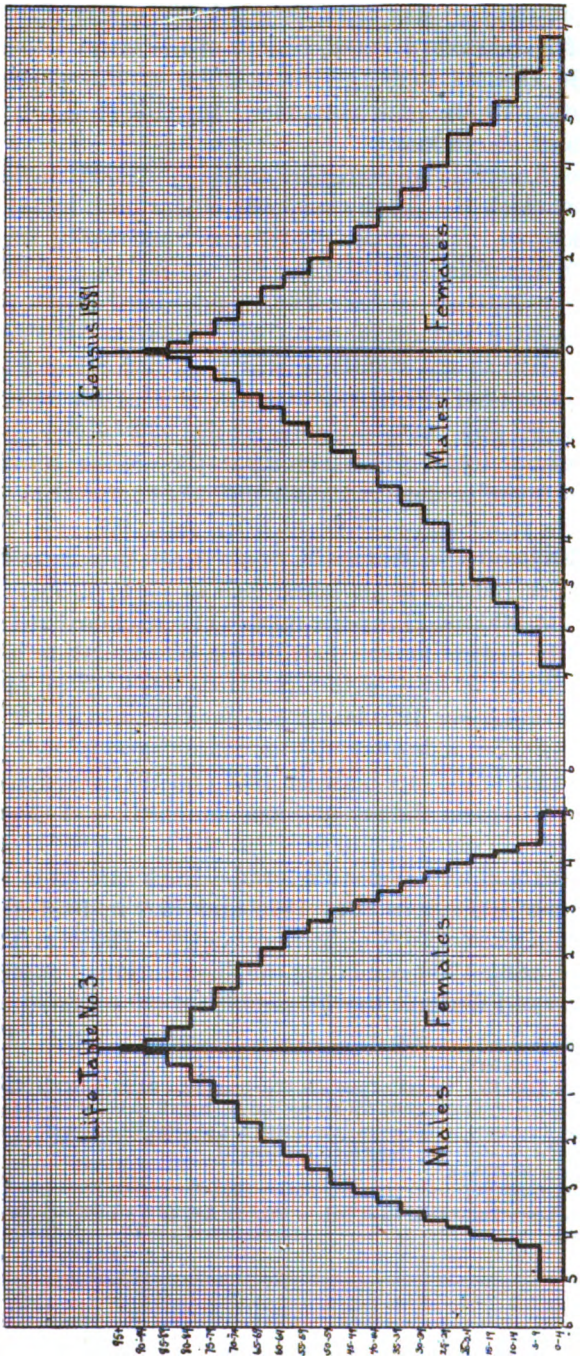


DIAGRAM II. STANDARD MILLION, ENGLISH LIFE TABLE NO. 3 AND ENGLAND AND WALES CENSUS OF 1881.

Digitized by Google

was 74.8 per 100,000, an increase of 15.9 per 100,000, or 27.0 per cent.* As 27.0 is about fifteen sixteenths of 28.7, it is evident that about one sixteenth of the increase in England's crude death rate from cancer in the decade 1891–1900 as compared with the decade 1881–1890 was due to the decennial change in the composition of the population.

d. In New York City the crude death rate from cancer was 65.5 per 100,000 in 1900 and 78.0 in 1910, an increase of 12.5 per 100,000, or 19.1 per cent.† The standardized death rate in 1910, accepting the distribution of population in New York City in 1900 as the standard, was 76.6, an increase of 11.1 per 100,000, or 16.9 per cent.‡ As 16.9 is about eight ninths of 19.1, it is evident that about one ninth of the increase in the crude death rate of New York City between 1900 and 1910 was due to the change in the composition of the population.

e. In London the crude death rate from cancer was 93.7 per 100,000 in 1901 and 107.2 in 1911, an increase of 13.5 per 100,000, or 14.4 per cent.§ The standardized death rate in 1911, accepting the distribution of population in London in 1901 as the standard, was 95.5, an increase of 1.6 per 100,000, or 1.7 per cent.|| As 1.7 is less than one eighth of 14.4, it is evident that more than seventh eighths of the increase in the crude death rate of London between 1901 and 1911 was due to the change in the composition of the population between those dates.

The general result of these calculations is to show that a considerable part of the increase in the crude death rate from cancer, a part ranging in these five cases from 5 to 88 per cent. and averaging about one third, is due to changes in the sex and age composition of the population. From the evidence I conclude that on the average not less than one tenth of the apparent increase¶ may be thus explained and that, wherever the birth rate as well as the death rate has been falling over a

* See Appendix, Table XVII.

† See United States Census Bureau, *Mortality Statistics, 1909*, p. 156, and *1910*, p. 103.

‡ See Appendix, Table XVIII.

§ *Hoffman*, p. 605, Table XVI.

|| See Appendix, Table XIX.

¶ I select this fraction somewhat arbitrarily because it comes near the result for the American registration states and the smaller of the two results for England. Results for a great country are likely to be nearer the average than those for any city, however populous.

Digitized by Google

period of years and in consequence the proportion of adult and aged persons in the population has been rapidly rising, there the change in the sex and age composition of the population may explain a much larger part of the increase in the crude death rate from cancer.

The article of Messrs. King and Newsholme, then, is faulty at a point where it has not hitherto been criticized, its unhesitating reliance upon standardized death rates of the usual type and its rejection of all material which does not lend itself to that method. It is valuable because of its emphasis upon the influence of improved diagnosis as a factor and even more significant because it pointed out that other causes competent to produce the result must be detected and measured by whoever denies that any part of the apparent increase of cancer is real.

Since their article was published a few other intensive studies have appeared, although none, I believe, indicates so firm a mastery of the method to be pursued and none carries out so thoroughly the distinction between accessible and inaccessible cases. The results of these may now be briefly summarized.

Massachusetts (1850–1895). In his first study Dr. Whitney concluded that "The increase is probably due to better diagnosis and registration and until the ratio of deaths over 30 years [sc. to all deaths at that period of life] has reached 8 to 9 per cent., which is shown by autopsies to be the true rate for cancer, it is not justifiable to speak of the increase as inherent in the disease itself."* But in a later article he admitted the existence of "a steady increase in the deaths [sc. death rates?] from cancer which cannot be explained by improved registration and more careful diagnosis."†

District of Columbia (1881-1900). Dr. Smith accepts the views of Messrs. King and Newsholme and believes that the evidence does not demonstrate any increase in the true death rate from cancer.‡

Providence, R. I. (1856–1905). Dr. Chapin thinks it prob-

* Wm. F. Whitney, "Alleged Increase of Cancer in Massachusetts," Shattuck Lecture before Massachusetts Medical Society, June 11, 1901, p. 20.

† "Statistics of Cancer in Massachusetts," in *Boston Med. and Surg. Journ., CLXII* (May 19, 1910), p. 673.

‡ C. A. Smith, "District of Columbia Cancer Record," in *Am. Med., III* (1902), pp. 916–19.

Digitized by Google

able, "that there has been no real increase in those forms of cancer which are visible or the diagnosis of which is easily made," namely, cancer of the breast and cancer of the uterus; and that "the apparent great increase in the mortality from cancer is due to improved diagnosis and changes in the sex and age distribution of the population."*

Breslau (*1876–1900*). Dr. Frief showed that the death rate from cancer of the breast and uterus had been stationary and the whole increase was due to an increase in cancer of the abdomen. He concluded that, when the needed precautions had been taken and the needed corrections made, no regular and decided increase in the death rate from cancer was demonstrable.†

Utrecht (*1872–1902*). Dr. Astro showed that the death rate from those forms of cancer which were easily diagnosed fell to three fourths of the initial rate, while the death rate from cancer of other forms more than doubled.‡

Amsterdam (*1862–1902*). Dr. Van Konijnenburg concluded that it was impossible to decide with confidence how much of the increase in the recorded death rate from cancer was due to a real increase and how much to improved diagnosis.§

Paris (*1886–1905*). Dr. Bertillon showed that, during these 20 years, the death rate from cancer of the mouth, breast, and female generative organs hardly increased at all, but that from cancer of the stomach increased rapidly and that from cancer of the intestines more rapidly. In other words, the increase in the recorded death rate was practically confined to cases of inaccessible cancer.‖

Netherlands (*1867–1909*). Director Methorst was unable to determine whether the increase of cancer was due to a greater prevalence of the disease or to greater accuracy in its diagnosis, the latter influence certainly accounting for a part of the increase.¶

* *City Registrar's Report, 1905*, pp. 89–93.

† *Klinisches Jahrbuch, XII* (1904), p. 154.

‡ Quoted in Van Konijnenburg, *Mortalité par Cancer à Amsterdam, 1862–1902*, pp. 14 and 11. Müller, Amsterdam, 1911.

§ *Ibid.*

‖ *Annuaire Statistique de la Ville de Paris*, XXV (1904), p. 202.

¶ International Statistical Institute, *Bulletin*, XIX, pt. III, pp. 310–312.

Digitized by Google

England and Wales. Dr. Stevenson in a careful and illuminating study of English cancer death rates,* published in 1913, reached no conclusion upon the main question. In the following year Major Greenwood, Jr., and Frances Wood, discussing Dr. Stevenson's evidence and arguments, reached the general conclusion "that there is some truth in the popular opinion that the real incidence of cancer has increased."†

Both of these English papers and no little current discussion seem to favor the view that the general question with which this article is concerned, Is the mortality from cancer increasing? can best be attacked by breaking it up into a series of subordinate questions like, Is cancer of the breast increasing? and answering the former only by finding the resultant of various and probably conflicting currents of change in cancer of various parts or organs. I cannot but feel that this is postponing attack upon the simpler statistical problem until a series of more difficult ones are solved. In other words, however unsatisfactory the present discrimination and reporting of cancer as a cause of death may be, and I believe them, as a rule, to be very unsatisfactory, they are probably much more accurate and trustworthy than the reporting of the organ or part of the body primarily affected by the disease.

These are all the significant intensive studies of the increase of cancer which have come to my attention.

An incidental result of rejecting as I do the method of standardized rates as the only or main method, is that it opens the door for that extensive study of the material in the recent international compilations to which we now turn our attention.

When inquiring how far, if at all, improvements in the diagnosis of cancer account for the increase in reported cancer mortality, we should recall the four ways in which statements of the cause on certificates of death may be obtained:‡ (1) from a medical examiner, authorized and required to perform an autopsy whenever necessary for confirming the diagnosis; (2) from an officer required to examine the body and consult with the attending physician but not authorized to perform

* *74th Annual Report of the Registrar-General*, for 1911.

† Royal Society of Medicine, *Proceedings*, VII, no. 1, pp. 119–170.

‡ *Cf.* von Mayr, *Statistik und Gesellschaftslehre*, II, §75 (Die Todesursachen), pp. 309–329; and Prinzing, *Medizinisches Statistik*, pt. 3, chap. 6. (Die Todesursachen), pp. 318–406.

Digitized by Google

an autopsy; (3) from the attending physician, midwife, next of kin, coroner, or person best informed; (4) from whoever reports the death to the registration officer.

When the figures are to be used for such a purpose as ours the fourth method is valueless. The first method would be most accurate but is nowhere in use at present. As the frequency of autopsies is connected with the increasing number and size of cities and with the development of hospital facilities,* this method may come into general use in the remote future. At present accuracy in the statement of the cause of death may be expected only for cities and countries in which the cause is regularly certified by a medical examiner or attending physician and even in such cases it depends upon the ability of the physician to diagnose and his willingness to certify cancer as a cause.

Cancer is seldom reported by a layman as a cause of death.† Hence the proportion of deaths ascribed to that disease usually depends upon the proportion of physicians to population and more especially upon the proportion of deaths reported by physicians to all deaths. Have these proportions been increasing?

The following figures‡ show that the proportion of physicians in Germany has increased:

Date.	Number of Physicians to 100,000 Population in Germany.
1876	32
1887	33
1898	44
1904	54

That a similar change was in progress in England and Wales down to 1901 appears from the following figures:

Date.	Number of Physicians to 100,000 Population in England and Wales.
1881	58
1891	66
1901	70
1911	65

* In 15 large cities of Switzerland 26 per cent. of the deaths in 1900 and in the city of Basel 34 per cent. of the deaths in 1902 were followed by autopsies. Of the autopsies in Basel 86 per cent. occurred in hospitals and only 14 per cent. outside. *Cf.* Prinzing, *op. cit.*, p. 320.

† *Cf.* pp. 740, *f.*

‡ Prinzing, *op. cit.*, p. 541.

Digitized by Google

There is no trend in the United States similar to that in England and Germany.

Date.	Number of Physicians to 100,000 Population in the United States.
1880	171
1890	166
1900	174
1910	164

The proportion of physicians to population in this country is about three times as great as in Germany and more than twice as great as in England, but is not increasing. Probably much of the difference is due to differences in the method of enumeration and to the readiness with which American enumerators accept without verification the claims of persons of questionable standing to be physicians. No doubt the quality, if not the relative number, of American physicians has improved within a generation and as a result their ability to diagnose cancer by clinical means, while still far from adequate, is probably greater than it was in 1880.

About the proportion of deaths in which the cause is certified by a physician, we have no satisfactory information for the United States or any part of it. But even if it has not increased during the last generation, the rise in the level of the profession—a change which occurred earlier in some European countries—has led to a similar result, for the incompetent physician is less likely than the well-trained one to certify cancer as a cause of death.

In several foreign countries, however, it is known that the proportion of deaths certified by physicians has increased of recent years In Saxony in 1873 only 37 per cent. but in 1906, 66 per cent. of the deaths were reported with a physician's statement giving the cause.* In Bavaria the proportion of decedents whose deaths were medically certified rose from 45 per cent. for the years 1845–1850 to 65 per cent. for the years 1901–1902. In Würtemberg it rose from 47 per cent. for

* Georg Radestock, "Zur Statistik der ärtzlichen Beglaubigung von Todesursachen," in *Zeitschrift des k. Sächsischen Statistischen Landesamtes*, LIV (1908), Pt. I, pp. 133–139.

Digitized by Google

1876–1881 to 63 per cent. for 1899–1900; in Baden, from 52 per cent. for 1852–1859 to 70 per cent. for 1892–1901. In Switzerland the statistics of causes of death are perhaps better organized than anywhere else. Dr. Prinzing awards it this preëminence among the European countries.* Particularly important for our purposes are the provisions in Switzerland whereby the attending physician makes a confidential report upon the cause of death to the statistical bureau and issues another certificate as a prerequisite to the burial. It may be more than a coincidence that the death rate from cancer in Switzerland is much higher than in any other country, but has not increased since 1900.† Both of these facts it would be easy to harmonize with the theory that the reported death rate from cancer depends largely upon the accuracy of certification and its increase upon the improvement of certification. The proportion of medically certified deaths in Switzerland increased from 82 per cent. in 1876–1880 to 96 per cent. in 1901–1903. In these 5 countries as a group, and they are the only ones for which I have found the information, the ratio of deaths medically certified to all deaths has increased about 0.5 per cent. yearly.

The probability that a physician was in attendance upon a decedent varies with the economic value of the endangered life and that depends not merely upon the wealth or position of the family but also upon the age of the individual Thus in Saxony in 1905 among children under 1 year of age less than 2 deaths in 5 were medically certified, but at ages 20–29 about 15 in 16 were so certified. With advancing age the proportion sank rapidly: at 50–59 it was 91 per cent.; at 60–69 it was 86 per cent.; at 70–79 it was 71 per cent.; and at ages above 80 it was 55 per cent.‡ Similar variations with age are found in the figures for Würtemberg.§

The death rate from cancer steadily increases with age until near the end of life.|| There is a large proportion of persons

* *Op. cit.*, pp. 321 and 325.

† Appendix, Table XX, and *Hoffman*, p. 674.

‡ Radestock, *op. cit.*, p. 137.

§ Prinzing, *op. cit.*, p. 322.

|| The highest death rate almost always occurs at some age period above 70. See Appendix, Tables XII, XIII, XV, XVI, XVII, XVIII, and XIX.

Digitized by Google

dying at an advanced age without a physician's care. Probably this proportion is decreasing with the progress of the years, accompanying the decrease in the proportion at all ages already demonstrated and the improved economic position of the poor. These facts together furnish good ground for believing that a steady and important shift in the certification of deaths from an indefinite cause, like old age, to cancer has been in progress and that this has been closely dependent upon a change in the authority stating the cause of death at high ages, *i. e.*, from layman to physician. The deaths in Saxony in 1905 were classified according to the cause assigned and the proportion under each cause which received medical certification was stated. At the top of the list was cancer, a cause certified by physicians in 97 per cent. of the cases, and near the bottom of the list was old age, a cause certified by physicians in only 42 per cent. of the cases. Hence the spread of medical certification would necessarily result in the appearance under cancer of not a few deaths which, under earlier conditions, would have been certified by a layman as due to old age.

Influences contributing to the same result are the extension of hospital facilities and the increase in the proportion of autopsies. It is difficult to get figures about either. The importance of autopsy in diagnosing cancer is illustrated by Riechelmann's statement that in 1895–1901 a municipal hospital in Berlin had 711 cases of cancer established by autopsy, 156 of which, or 22 per cent., had not been so diagnosed during the lifetime of the patient, although the diagnosis had been made with the greatest care by the highest authorities, thus excluding the carelessness with which the ordinary physician in Berlin still reported the causes of deaths occurring in his practice.*

To be sure, not all the mistakes are on the same side. Sometimes a clinical diagnosis of cancer is negatived by the findings at the autopsy. But the additions resulting from autopsy exceed the subtractions. This is indicated by Dr. Bashford's evidence. In 1909 he reported upon 9,488 fatal cases in hospitals which had been treated during the life of the patient as

* W. Riechelmann, "Eine Krebsstatistik vom pathologisch-anatomischen Standpunkt," in *Berliner Klinische Wochenschrift*, XXXIX (August 4, 1902), p. 729.

Digitized by Google

cancer and after death had been checked by autopsies. This check showed that 757 cases, or 8.0 per cent., of those so diagnosed were not cancer. But during the same period and in the same hospitals, 1,801 other cases of fatal disease, more than twice as many, which during life had not been diagnosed as cancer were shown by autopsies to be so. The net result was to show that the true number of deaths from cancer in these hospitals was 10,532 instead of 9,488, so that the number diagnosed as cancer, rightly or wrongly, by clinical means during the lifetime of the patient was only 90 per cent. of the deaths actually due to this disease. In cases of cancer attacking accessible parts of the body the number of deaths determined by clinical examination was 98 per cent. of the true number; but where inaccessible parts of the body were primarily involved the number of deaths determined by clinical examination was only 68 per cent. of the true number.*

Dr. Williams has suggested that "these were specially selected lots of difficult hospital cases and cannot be regarded as in any way typical of every day experience."† That is a question upon which a layman must speak with diffidence, but in my opinion the greater attention and skill bestowed upon hospital patients would more than offset the greater difficulty of diagnosis, if it exists.

If physicians attending hospital patients often fail to diagnose cancer, such failures must be far more common among physicians in general practice, in which case an increase in the ratio of deaths in hospitals to all deaths would mean an increase in the ratio of deaths ascribed to cancer to all deaths from it. This suggests the question, Is the ratio of deaths in hospitals to all deaths increasing? The only evidence for the United States is indirect. At the end of 1904, to every 100,000 people there were 87 patients in American public hospitals; at the end of 1910 there were 118, an increase of more than one third in 6 years. During the year 1904, the number of patients treated in public hospitals was 1,309 per 100,000; during the year 1910, it was 2,124, an increase of nearly two thirds in 6 years.‡

* *Lancet*, 1909, II (Sept. 4), p. 694.

† W. R. Williams, *The Natural History of Cancer*, p. 54.

‡ United States Census Bureau, *Benevolent Institutions, 1904*, p. 32; *Benevolent Institutions, 1910*, pp. 48–49.

Digitized by Google

In Greater New York the proportion of deaths in hospitals to total deaths has doubled in the 17 years since the enlarged city was organized, rising from 16.3 per cent. in 1898 to 32.9 per cent. in 1915. In old New York the same change was in progress through an earlier period, the proportion of deaths in hospitals to all deaths rising from 15.4 per cent. in 1889 to 22.0 per cent. in 1897.*

German figures giving the number of beds in general hospitals, public and private, indicate a change almost as rapid. In that Empire in 1877 there were 166 hospital beds to 100,000 people and in 1910 there were 400, nearly two and one half times the earlier ratio.†

The figures for Austria‡ point to a slower increase. Owing to differences in the classification of hospitals in the two countries, the following ratios are not comparable with those for Germany.

Year.	Hospital Beds per 100,000 Population.
1870	118
1880	124
1890	142
1900	179

In France the average annual number of beds in hospitals in the periods 1871–1880 and 1901–1905§ may be compared with the population in the middle of each period. The results show an increase in beds per 100,000 people from 193 to 236 in about 30 years. Probably a similar change is under way in nearly every civilized country and there is little doubt that it has materially increased the chance of correctly diagnosing cancer as a cause of death.

The preceding evidence shows that in every European country for which the information was available the proportion of physicians to population and the proportion of deaths certified by physicians have increased; that in the United

* See Appendix, Tables XXI and XXII.

† See Appendix, Table XXIII.

‡ Prinsing, *op. cit.*, p. 545.

§ From *Stat. annuelle des Instit. d'Assistance* (1906), cited in Webb, *New Dict. of Statistics*, 1911, p. 298.

Digitized by Google

States, although neither is known to have risen, the average competence of physicians has probably increased; that in several European countries even now a large proportion of deaths at ages when cancer mortality is highest are not medically certified; that cancer is the cause of death least often stated by a layman; and that hospital facilities are more numerous and autopsies more frequent than ever before. With this evidence in mind, we are in a better position to weigh the arguments tending to show that improvements in the diagnosis and certification of cancer, which have long been in progress and are still continuing, explain at least in large part the increase in its reported mortality.

Let us assume for a moment the extremely skeptical position that the entire increase in the reported death rate, aside from that part due to increasing longevity, is due to improved diagnosis; or that the only influence affecting the death rate from cancer which has been statistically demonstrated is age. In other words, assume that there are no demonstrated differences in the true cancer death rate of men and women, of white and colored, of residents in city and country or in different countries, but that all differences in the recorded rates of these groups at the same age are due to differences in the accuracy of diagnosis. From this assumption certain corollaries would follow. If they are shown to conform to the facts, the assumption is thereby and to that degree made probable.

1. Is the proportion of deaths reported as due to an unknown or ill-defined cause decreasing? In the part of the United States for which significant figures exist, the registration states of 1900, the number of deaths ascribed to ill-defined or unknown causes* since 1900 has been steadily diminishing. In 1900 there were 13,116 deaths from this group of causes; in 1915 there were 1,090, a decrease of 12,026, or 91.7 per cent. But these deaths are distributed to the several ages of life in somewhat the same proportion as other causes of death, while cancer is characteristically a disease of adults and aged persons. Hence it is fairer to exclude all deaths ascribed to ill-

* In general terms this group includes every death which was said to be due to ill-defined organic disease, sudden death, or a cause ill-defined or not specified.

Digitized by Google

defined or unknown causes in which the decedents were less than 30 years of age. In 1900 there were in these registration states 6,362 deaths of persons over 30 years of age which were ascribed to ill-defined or unknown causes; in 1915 there were only 520, or about one twelfth as many. As the population increased, while the deaths decreased, the death rate from such causes fell still faster, from 74.9 per 100,000 population over 30 years of age in 1900 to 4.5 per 100,000 in 1915, or one sixteenth of the earlier rate.

An estimate of the number transferred from ill-defined causes to all other causes may be reached in the following way. If the rate in 1915 had been the same as in 1900, the deaths above 30 years of age from unknown causes would have been 8,696. The difference between that number and 520, or 8,176, represents the number probably transferred from ill-defined and unknown causes to some specific cause. If we assume that these were transferred in a random way, or in the proportions which each of those causes bore to all causes except the ill-defined group, then 9.6 per cent. of the 8,176 transfers, or 784, went to cancer. The actual increase in the cancer death rate between 1900 and 1915 was from 144.3 to 202.3 per 100,000 population over 30 years of age, or 40.2 per cent. If these 784 estimated transfers were excluded, the increase would be only from 144.3 to 195.5, or 35.5 per cent. On these assumptions between one ninth and one eighth, 11.7 per cent., of the apparent increase was due to transfers from ill-defined causes to cancer. But in assigning causes of death the modern tendency is away from vague, indefinite causes to specific, definite causes. Hence, the gain by transfers to cancer has probably been greater than a mere chance distribution of the transferred cases would indicate. If so, more than one ninth of the increase in the cancer death rate of the registration states of 1900 between 1900 and 1915 was due to a certification of deaths in 1915 as due to cancer which, under conditions existing in 1900, would have been returned as due to ill-defined or unknown causes.

In dealing with this aspect of the question, if errors are to be avoided, the specific conditions surrounding certification and registration in each country must be known, and espe-

Digitized by Google

cially whether the residual class of ill-defined and unknown has retained the same meaning throughout. This knowledge I have found it difficult to acquire and for that reason I have made no study of such transfers in other countries.

But after this warning it may be pointed out that in England and Wales, according to *March*, the deaths ascribed to unknown causes fell from 45,273 in 1858 to 3,318 in 1910, a decrease of 93 per cent. However, in this case there is reason to suppose that the meaning of the term *unknown* has changed.* In Scotland between 1855 and 1910 the deaths from unknown causes fell off by three fifths and, as the population increased, the death rate fell off by three fourths; in Ireland the death rate from unknown causes fell by one third between 1881 and 1910. In Prussia between 1875 and 1910 the deaths from unknown causes decreased from 101,633 to 18,434, or nearly five sixths and the rate about eight ninths; in Baden the decrease in deaths from unknown causes was from 1,381 in 1881 to 17 in 1910; in Italy from 22,977 in 1890 to 3,455 in 1910 and in the more populous divisions of Australia since 1891 similar changes have been in progress with great rapidity. All this evidence seems to warrant us in concluding that in many countries the assignment of deaths to cancer, which under earlier conditions would have been assigned to unknown or ill-defined causes, has played an important part in exaggerating the increase in reported cancer mortality.

2. Is the proportion of deaths reported as due to tumor decreasing? The reported death rate from tumor, or, more exactly, benign tumor, has been decreasing, while the death rate from cancer, sometimes called malignant tumor, has been increasing. These changes in opposite directions are probably to be ascribed in large part, if not entirely, to a more accurate discrimination between cancer and tumor, or a greater willingness among friends of the decedent to have the word *cancer* appear on the death certificate. As a result many deaths in recent years have appeared in the cancer column which under earlier conditions would have appeared as tumor. American evidence regarding this change goes back to a period

* *March*, I, p. 829 and II, p. 235. But see the divergent figures for 1901 under this head in the two volumes.

Digitized by Google

long before 1900. In the census of 1880, this statement appears: "The reported number of deaths from cancer . . . is undoubtedly too small, if we use the term 'cancer' in the ordinary sense, for there is very little doubt that most of the deaths reported as due to tumor . . . might properly have been classed with cancer."* Ten years later the same authority said: "As it is impossible from the census data to distinguish the cases of tumor from those of cancer, they will be considered together."† The same treatment was accorded to them in 1900.‡ The diminishing proportion that deaths ascribed to tumor made of all ascribed to cancer and tumor combined is shown below.§

PER CENT. THAT DEATHS FROM TUMOR MAKE OF TOTAL FROM CANCER AND TUMOR.

Date.	Census Returns for all United States.
1880	12.0
1890	11.7
1900	10.4
	Returns for Expanding Registration Area of United States.
1900–1904	2.3
1905–1909	1.5
1910–1914	0.9

The first three lines refer to deaths from any part of the United States as returned by the enumerators outside, and by medical certification inside, of the registration area. The other three refer to the registration area alone, which embraced in 1900 two fifths and in 1914 two thirds of the population of the country.‖ This expansion has carried registration into districts where the diagnosis of cancer is less trustworthy, probably, and hence the preceding figures obscure the results of improved diagnosis in an unchanging area. An idea of the exaggerated emphasis upon tumor as a cause of

* U. S. Census Office, *Tenth Census*, XII, p. cxiv.
† U. S. Census Office, *Eleventh Census, Vital Statistics*, pt. I, p. 336.
‡ U. S. Census Office, *Twelfth Census*, III, p. clxxxi.
§ See Appendix Table XXIV, for fuller data.
‖ U. S. Bureau of the Census, *Mortality Statistics, 1914*, p. 9.

Digitized by Google

death in the enumerators' returns alone may be derived from the complete data,* by subtraction. It appears that outside of the registration area of 1900, and thus in districts where the only source of information was enumerators' returns, 22 per cent. of the deaths from cancer and tumor were ascribed to tumor, but within the registration area at the same date only 2.9 per cent. of the deaths in that group were ascribed to tumor.

The evidence derived from the registration states of 1900 for the period 1900 to 1915 points to the same conclusion and is more weighty because the area covered by the figures did not change. In this case it is safe to speak of rates rather than of proportions. The results are as follows:

DEATH RATE PER 100,000 POPULATION IN REGISTRATION STATES OF 1900. (a)

Year.	Cancer.	Tumor.	Both.	Per Cent. of Deaths from Cancer.
1900	63.9	2.0	65.9	97.0
1901	65.8	1.6	67.4	97.6
1902	65.6	1.4	67.0	97.9
1903	69.0	1.7	70.7	97.6
1904	70.4	1.4	71.8	98.1
1905	72.4	1.4	73.8	98.1
1906	73.1	1.3	74.4	98.3
1907	75.7	1.2	76.9	98.5
1908	76.8	1.0	77.8	98.7
1909	79.4	1.0	80.4	98.8
1910	83.0	1.2	84.2	98.6
1911	83.9	0.8	84.7	99.1
1912	85.9	0.7	86.6	99.2
1913	88.7	0.5	89.2	99.4
1914	89.1	0.6	89.7	99.3
1915	91.8	0.5	92.3	99.4

(a) For the full table see Appendix, Table XXV.

In these 15 years the death rate from cancer has increased 27.9 per 100,000 population, but the death rate from tumor has fallen 1.5 per 100,000, so that the death rate from both combined has increased only 26.4 per 100,000. Thus the transfer from tumor to cancer seems to account for more than one twentieth of the total increase.

The influence of this progressive transfer of cases from tumor to cancer through a longer period is clearly revealed in the Massachusetts figures.† As the state classification differs

*Appendix, Table XXIV.

†For the full table see Appendix, Table XXVI.

Digitized by Google

from the federal and the latter is available only since 1900, I have relied throughout upon the former.

DEATH RATE PER 100,000 POPULATION IN MASSACHUSETTS.

Period.	Cancer.	Tumor.	Cancer and Tumor.	Per Cent. of Deaths from Cancer.
1853–1857	19.3	5.3	24.6	78.5
1858–1862	25.8	4.9	30.7	83.9
1863–1867	29.7	5.3	35.0	84.9
1868–1872	34.9	5.8	40.7	85.8
1873–1877	37.4	5.4	42.8	87.3
1878–1882	50.8	3.8	54.6	93.0
1883–1887	56.1	3.3	59.4	94.5
1888–1892	60.6	3.1	63.7	95.1
1893–1897	67.1	3.8	70.9	94.6
1898–1902	71.0	3.0	74.0	96.0
1903–1907	83.3	2.5	85.8	97.1
1908–1912	90.2	1.2	91.4	98.6
Increase or decrease (—) per 100,000	70.9	—4.1	66.8	

These figures indicate that about one seventeenth of the increase in the cancer death rate of Massachusetts was due to the certification of deaths at the close of the period as caused by cancer which under earlier conditions would have been ascribed to tumor.

The only article I have met with in which deaths from cancer and tumor are separately tabulated is one dealing with the statistics of Breslau for the period 1876–1899.* From its figures I have compiled the following:

Period.	Deaths from		Per Cent. from Tumor.
	Cancer and Tumor.	Tumor.	
1876–1879	666	94	14.1
1880–1884	969	67	6.9
1885–1889	1,291	69	5.4
1890–1894	2,024	204	10.1
1895–1899	2,729	307	11.3

A comparison of these figures with the others suggests that many of the deaths in Breslau ascribed to tumor are probably due to cancer and that there has been no steady improvement

* F. Frief, "Die in den Jahren 1876–1900 in Breslau vorgekommenen Todesfälle an Krebs," in *Klinisches Jahrbuch*, XII (1904), pp. 133–200.

Digitized by Google

in discriminating between the two causes of death such as is indicated by the American figures.

In England and Wales the Registrar-General in 1884 began the practice of sending an inquiry to every physician who reported a death without a satisfactory statement of its cause, and in 1887 extended this procedure to include every case of a death certified as due to tumor. In 1886 the deaths reported as due to cancer were 16,243, in 1911 they were 35,902, an increase of 18,659 in 25 years. But in 1911 there were 656 deaths which had been at first reported as due to tumor and which, after correspondence, were transferred to cancer. This number equals 3.5 per cent. of the increase. Probably a large, perhaps an equally large, number of cases in 1911 were returned in the first instance as cancer, although under the conditions of greater carelessness or ignorance prevalent before 1887 they would have been reported as tumor. This English evidence seems to corroborate the American by indicating that a significant part of the increase in the cancer death rate of recent years, a part which may be conservatively estimated as one twentieth, has been due to a transfer of cases from tumor to cancer.

3. Is the proportion of deaths reported as due to old age decreasing?

In the American registration states of 1900 the number of persons reported as dying of old age is steadily and rapidly decreasing. In 1900 it was 10,015 and in the same states in 1915 it was 4,380.* This change is not due to a decrease either in the total population or in those at high ages; for both the population and the proportion who die late in life to all deaths are growing. It is due rather to the unwillingness of many recording officials to receive vague statements of the cause of death and their efforts to secure on the death certificate a statement more specific and exact than "old age." Thus, in the *Manual of the International Classification of Causes of Death*, the *vade mecum* of American registration officers, the instruction under "old age" reads: "A very indefinite and unsatisfactory return as a cause of death. It is usually due to carelessness on the part of the physician and a more specific

* Appendix, Table XXVII.

Digitized by Google

statement should be required." The decrease in the deaths ascribed to "old age" is due, then, to the increasing care with which returns are made and the closer supervision of them by the registration officials, and the change is a rough index of the improvement of registration. Doubtless many of the deaths which would formerly have been ascribed to "old age" now appear under cancer, contributing to swell the increase in the recorded, as distinguished from the actual, cancer death rate.

To estimate the proportion of the increase in cancer mortality which can be thus explained, it is necessary to estimate first the number of deaths transferred from old age to all other causes and then the proportion and number which have gone to cancer. For this it is probably best to assume a constant ratio between the number of deaths ascribed to old age and the number of deaths of all persons above 60 years of age. In the registration states of 1900 this proportion in 1900 was 9.9 per cent. and in 1915 it was 3.2 per cent. If the per cent. had remained unchanged, 13,697 deaths would have been ascribed to old age in 1915; the number actually ascribed to it was 4,380. Hence it is probable that there were not far from 9,317 deaths in 1915 which under earlier conditions would have been ascribed to old age but which were ascribed to some other cause as a result of improvements in reporting causes of death. Of these 9,317 deaths, the number transferred to cancer may be estimated by assuming that the proportion was equal to that which the deaths above 60 years of age from cancer make of the deaths at that time of life from all causes except old age. This would indicate that 9.3 per cent. of the 9,317 deaths, or 866, were transferred from old age to cancer. If these 866 deaths be subtracted from the total ascribed to cancer in 1915, the increase in the cancer death rate during the 15 years 1900–1915 is reduced from 27.9 to 24.6 per 100,000. In other words, the transfer from old age to cancer serves to account for about 12 per cent. of the entire increase. But the tendency to change from old age to cancer must have been much greater than the tendency to change from old age to heart failure or general debility or some other meaningless term. If so, the transfers from old age to cancer in this

Digitized by Google

instance must account for more than 12 per cent. of the increase in cancer mortality.

The difficulties which prevented an international study of the transfers from ill-defined and unknown causes to cancer do not equally hamper a study of the transfers from old age. I have found no reason to suppose that the heading *old age* or its counterpart in the international classification, *senile debility* or *senility*, has materially changed its meaning. The decreasing death rate from this cause is shown in the following table, giving the death rate from old age for every fifth year since the record began. The data are taken from *March*.

DEATH RATE PER 100,000 POPULATION FROM OLD AGE, OR SENILITY, VARIOUS EUROPEAN STATES.

Year	England and Wales.	Scotland.	Ireland.	Norway.	Switzerland.	Prussia.	Bavaria.	Baden.	Italy.
1855		191							
1860	143	215							
1865	126	183	173						
1870	128	192	295	38					
1875	119	169	405	49		264			
1880	100	129	432	48	112	251			
1885	103	123	381	45	125	247		197	
1890	98	116	401	44	130	248	245	215	101
1895	98	110	388	107	128	221	213	182	125
1900	99	108	364	175	170	236	209	173	148
1905	92	84	197	194	73	201	191	154	143
1910	88	74	208	187	55	167	161	146	140

The preceding figures show that in most of the countries the death rate from old age is now falling and that in several it has fallen without interruption from the first. Probably in all these countries the proportion of deaths at ages above 60, 70, or 80 to all deaths or to all persons living is increasing. Hence many deaths must be now ascribed to other causes which, under earlier conditions, would have been ascribed to old age, and no small proportion of these must have gone to cancer.

The preceding estimates explain away not less than three tenths of the increase in the reported mortality from cancer, namely, transfers from ill-defined or unknown causes to cancer perhaps 12 per cent., transfers from old age to cancer perhaps 12 per cent. and transfers from tumor to cancer perhaps

Digitized by Google

6 per cent. But the two larger figures have been reached by assuming that the transfers from indefinite to definite causes had a random distribution among the groups to which they went, while in fact the distribution tends to center on specific and precise causes. For this reason the results may fairly be deemed minimum figures and not a little below the truth. Accordingly, I conclude that from three tenths to one half of the increase in the death rate from cancer in the registration states of 1900 is due to the causes mentioned and is apparent rather than real.

If to these results we add the earlier conclusion that, on the average, not less than one tenth of the increase in cancer mortality is due to the increasing proportion of adult and elderly persons in the population, I conclude that not less than four tenths and perhaps as much as six tenths of the increase in cancer mortality has been explained as only apparent.

These are all the results of improved diagnosis and certification upon the recorded cancer mortality which I have found any means of measuring even roughly. There are other changes, however, which have already been mentioned and need examination, although their influence cannot be measured.

4. Is the mortality from cancer low where there are few physicians or where a small proportion of the death certificates are prepared by physicians and vice versa?

We have already (p. 725) touched upon the relation between the reported death rate from cancer and the proportion of deaths certified by laymen. But additional evidence on this point may now be presented. In Saxony, of all deaths certified as due to cancer the per cent. certified by laymen was 18 in 1876 and steadily decreased to 4 in 1903.* Hence, where certificates from doctors are few, the cancer mortality is likely to be low. In this respect no country, perhaps, includes greater extremes than Austria; Vienna is one of the medical centers of the world; Bukowina and Galicia are most meagerly furnished with physicians. *Prinzing* gives for each of the 17 provinces of Austria in the five year period 1896–

* Georg Radestock, "Die Krebsterblichkeit im Königreich Sachsen in den Jahren 1873–bis 1903," in *Zeitschrift des k. Sächsischen Statistischen Landesamtes*, LI, Pt. II, p. 263.

Digitized by Google

1900 the per cent. of deaths the causes of which were certified by physicians,* and *Hoffman* gives the death rates from cancer, 1907–1911, in the same provinces.† As the two returns do not speak for the same period, I have sought for the per cent. of medically certified deaths in the later years. The source‡ gives the information only for 1907 and 1910. From this material I have computed the per cents. in the third column of the following table:

Province.	Death Rate from Cancer per 100,000 Population 1907–1911.	Per Cent. of Deaths which were Medically Certified	
		1896–1900.	1907 and 1910.
Salzburg	140	99.0	99.9
Upper Austria	136	98.2	99.3
Vorarlberg	125	96.3	98.4
Lower Austria	120	99.7	99.9
Tyrol	110	93.0	94.2
Bohemia	104	99.1	99.7
Trieste	99	98.8	99.9
Moravia	97	94.4	95.8
Styria	96	79.9	83.7
Corinthia	83	72.4	77.5
Silesia	67	70.1	74.7
Carniola	49	35.0	35.9
Görz and Gradiska	46	43.1	47.1
Bukowina	46	32.8	33.7
Istria	43	44.7	51.0
Galicia	33	26.5	27.6
Dalmatia	24	29.1	30.5
All Austria	79	68.0	68.5

Among the many factors which influence the reported death rate from cancer in Austria, there is one, the proportion of deaths certified by physicians, so potent as largely and, where the differences are great, almost exclusively to determine the rate. Evidently the very low reported death rate from cancer in Dalmatia and Galicia is due to the few reports from physicians and not to the infrequent occurrence of cancer in those provinces. A comparison of the second and third columns in the table shows that the proportion of medically certified deaths in every province increased between the earlier and later dates. The increase in the Austrian death rate from

* P. 326.

† P. 680.

‡ *Oesterreichische Statistik.* The figures were kindly furnished by the New York Public Library.

Digitized by Google

cancer during the same period from 69 to 78 per 100,000 was doubtless connected with this spread of medical certification.

The following table has been compiled from one in *Hoffman** by arranging the countries of Europe in the order of the cancer death rate.

Country.	Period.	Death Rate from Cancer per 100,000 Population.
Denmark (cities)	1908–1912	142
Switzerland	1906–1912	124
Holland	1908–1912	106
Sweden (cities)	1907–1911	105
Scotland	1908–1912	103
England and Wales	1908–1912	98
Norway	1908–1912	97
German Empire	1908–1912	87
Ireland	1906–1912	81
Austria	1908–1912	80
France	1906–1910	76
Belgium	1908–1912	67
Italy	1908–1912	65
Rumania (cities)	1907–1911	62
Spain	1908–1912	52
Greece (cities)	1904–1908	52
Hungary	1908–1912	46
Portugal	1906–1910	22
Serbia	1905–1909	12

This suggests that in the countries of Europe the cancer death rates vary, as in the provinces of Austria they have been shown to do, in general agreement with the number and quality of the physicians, or the proportion of deaths which are medically certified. But I have found no figures showing the proportion of physicians to population or the proportion of medically certified deaths in these countries and therefore must content myself with raising the question without answering it.

5. Can any differences be traced between the amount or the increase of cancer mortality in city and country?

The cities of the United States having at least 25,000 inhabitants in 1900 had 232 physicians to every 100,000 people and the rest of the country had 153, or only two thirds as many. Probably the number of visits made by a physician annually is greater in cities than in the country and the quality of his training is better. The proportion of young adults with lives

* P. 594.

Digitized by Google

economically of high value is also greater in cities, and not a few invalids migrate to cities to avail themselves of medical and hospital advantages. All these circumstances would lead one to look for a higher cancer death rate in cities, if the assumption were correct that a direct and important connection exists between the number and quality of the physicians in a community, and its reported death rate from cancer. On the other hand, in country districts the proportion of aged persons peculiarly liable to cancer is larger. If, notwithstanding this physical basis for a higher cancer death rate in the country, the cities actually have the higher reported rate, then the difference is doubtless due in considerable measure to the larger number and greater ability of city physicians, if we assume that there is no actual difference in the liability of the two populations at the same ages to die of the disease. Now, in the cities of the registration states of 1900 the death rate from cancer in 12 of the 16 years during the period 1900–1915 was greater than in the country districts, indicating that in the United States there may be a connection between the proportion of physicians to population and the reported death rate from cancer.*

Between 1900 and 1915 the death rate from cancer in the cities of the registration states increased from 66.0 to 91.3 per 100,000, or 38.0 per cent. Meanwhile the country rate increased from 61.4 to 92.8, or 51.0 per cent., a more rapid increase in the country rate. During the same period we may be sure from evidence already given that hospital facilities and their use increased quite out of proportion to population. Hence there can be little doubt that deaths from cancer occurring in a hospital are now a larger proportion of all such deaths than formerly they were. And it is almost as certain that the current of cancer patients from country districts to city hospitals is increasing. If so, and if there are no other real differences between the cancer mortality of city and country districts than those due to differences in the proportion of elderly persons, the more rapid increase of the reported death rate from cancer in the country shows that some change, perhaps improved diagnosis and certification, has progressed so

* Appendix, Table XXVIII.

Digitized by Google

much more rapidly in the country as to overbalance the growing drift of cancer patients from country to city.

The tables in the international publications already mentioned facilitate a comparison between cities and country districts in certain European countries. Thus for England the population and deaths from cancer since 1880 are given for Birmingham, Bradford, Derby, London, and Liverpool by *Falkenburg*, and for the whole country by *Hoffman*. By subtracting the urban totals from those for all England a rough division between city and country is reached. In the same way the rates for Edinburgh and Glasgow combined may be compared with those for the rest of Scotland, those for Christiania with those for the rest of Norway, those for Amsterdam, the Hague, Rotterdam, and Utrecht with those for the rest of the Netherlands, and those for Vienna and Prague with those for the rest of Austria.* The computation shows that in sixteen out of twenty cases the crude urban rate was higher than that in the rest of the country and that in every country but England the increase in rate between 1881 and 1909, or in the case of Austria 1881 and 1908, was more rapid in the country districts, thus indicating that in Europe the difference between city and country has been decreasing. The figures for Austria are especially noteworthy. In 1881 the death rate from cancer in Vienna and Prague was 119.0, in 1908 it was 133.0. In the rest of Austria the rate in 1881 was 38.6, in 1908 it was 75.0. While the urban rate was increasing 12 per cent., the rural rate increased 94 per cent. In view of the other evidence which has been presented regarding conditions in the various provinces of Austria, it seems probable that nearly all of this increase in cancer mortality outside of the two capitals is due to improved diagnosis and certification.

Apparently, then, cancer mortality in cities, or at least in great cities, is usually higher but increasing more slowly than it is in country districts, so that in this regard the difference between city and country is decreasing. This is especially true wherever the initial differences between city and country are very wide.

6. Is the death rate from cancer among Negroes less than it is among whites?

* For detailed figures, see Appendix, Table XXIX.

Digitized by Google

The greater longevity of the whites would result in a higher cancer death rate among all whites, even if there were no difference in the death rates of persons of the same age belonging to the two races. If the true cancer death rates of the two races at the same age periods are the same but medical attendance and certification are poorer among Negroes, then the reported Negro death rate from cancer would be the lower. Unfortunately the death rates by race are available for only 6 years, 1910–1915. They speak for the registration states of 1910 and embrace an annual average of over 39,000 deaths from cancer of whites and 900 deaths of colored, nearly all of them Negroes. The results are as follows:*

Date.	Death Rates from Cancer per 100,000 Population Classified by Race in Registration States of 1910.		Per Cent. that Death Rate among Colored is of that Among White=100.
	White.	Colored.	
1910	76.8	52.6	68.4
1911	77.0	59.3	77.0
1912	79.8	59.3	74.3
1913	82.6	64.0	77.5
1914	83.5	65.4	78.3
1915	85.4	68.8	80.6
Per Cent. increase.	11.2	30.8	

The reported death rate of colored from cancer is from two thirds to four fifths that of the white, but is increasing faster, so that the difference between the races is diminishing. In view of the short period, the small number of Negro deaths involved and the erratic figures, no great weight can be given to the results.

I have obtained similar figures for the registration states of 1900, a much smaller area, in the same years. They have one advantage—they can be safely compared with figures for 1900.†

* For figures in full see Appendix, Table XXX.

† The figures for 1900 refer to the census year rather than to the calendar year and the area embraced differs from that of the registration states of 1900 by excluding Indiana. But these differences do not materially affect the significance of the comparison here made.

Digitized by Google

Date.	Death Rates from Cancer per 100,000 Population Classified by Race in Registration States of 1900.		Per Cent. that Death Rate among Colored is of that Among White=100.
	White.	Colored.	
1900	62.5	41.3	66.1
1910	83.4	60.5	72.5
1911	84.1	74.2	88.5
1912	86.2	71.5	83.0
1913	89.0	70.8	79.6
1914	89.4	72.1	80.6
1915	92.1	79.4	86.2
Per Cent. increase.	47.4	92.2	

Here again it appears that the death rate from cancer among the colored is lower but increasing faster than that among the white. The changes from year to year are irregular. A comparison of the two tables shows that the death rate from cancer is uniformly greater for each race in the smaller region from which the states recently added to the registration area are excluded, as in the second table, a fact agreeing well with the theory that differences in the cancer death rates are due largely to the efficiency of the registration system, which even under the best of fostering care must be a plant of slow growth.

Mr. Hoffman's tables* show that in 8 large Southern cities between 1891 and 1914, the first and last years covered, cancer mortality among whites increased by 73 per cent. and that among Negroes by 92 per cent. He has long given special attention to the statistics of Negroes in the United States and notes that prior to the Civil War cancer among them was rare, that the present difference between the two races is less pronounced, but that the cancer mortality of Negroes is still materially below that of whites.† These conclusions by one who accepts the opinion of qualified medical observers that "cancer is exceptionally rare among primitive peoples,"‡ are equally consistent with my own belief that much, if not all, of the difference is due to differences in the accuracy of diagnosis.

* *Op. cit.*, p. 474.
† *Ibid.*, p. 15.
‡ *Ibid.*, p. 147.

Digitized by Google

7. Has the death rate from cancer increased more rapidly at the higher age periods than earlier in life?

The Census Bureau has kindly furnished me with a table showing the cancer mortality in the registration states of 1900 classified by age and sex for the two years 1900 and 1910.* From these figures the following percentages of increase have been computed.

PER CENT. INCREASE IN THE DEATH RATE FROM CANCER BY SEX AND AGE IN THE REGISTRATION STATES OF 1900 FOR THE PERIOD 1900–1910.

Age.	Per Cent. Increase.	
	Males.	Females.
25–34	4	16
35–44	17	13
45–54	28	17
55–64	44	24
65–74	31	33
75+	34	42

The figures are not perfectly regular and yet they do show clearly that in each sex the increase at the higher ages is more rapid than that at the lower ages.

Similar evidence extending over a longer period of time is furnished by the Massachusetts records. The rates for a single year, when classified by sex and age, are often based on too few cases to be entirely trustworthy. The records cover the 6 census years 1860–1910, which, with due consideration of the sex differences, gives 10 cases.† In one of the 10 the most rapid increase was at ages over 80; in 3 it was at 70–79, in 3 it was at 60–69, and in only 3 was it at any period below 60, all 3 being at 30–39. This supports the previous conclusion that the increase of cancer mortality is greatest toward the end of a normal lifetime.

Similar rates for a series of decades are to be found in the successive decennial supplements to the *Reports of the Registrar General* and for 1901–1910 in his *Seventy-third Annual Report* for 1910. From them the following percentages of increase have been derived. As each rate speaks for a 10

* Appendix, Table XXXI.
† Appendix, Table XXXII.

Digitized by Google

year period and for a large population, the results are more regular and more significant than those previously reached.

PER CENT. OF DECENNIAL INCREASE IN THE DEATH RATE FROM CANCER BY SEX AND AGE IN ENGLAND AND WALES.

Age.	Per Cent. of Decennial Increase.			
	1861–1870. to 1871–1880.	1871–1880 to 1881–1890.	1881–1890 to 1891–1900.	1891–1900 to 1901–1910.
	Males.			
25–34	18	14	26	?
35–44	17	25	29	13
45–54	32	42	30	25
55–64	33	45	38	28
65–74	74	44	42	34
75+	32	31	49	?
	Females.			
25–34	9	0	2	?
35–44	18	8	5	−1
45–54	16	16	14	5
55–64	20	22	22	11
65–74	26	28	29	22
75+	26	31	39	?

The preceding figures all show that the increase in cancer mortality is greater at the higher ages. Incidentally the last series shows also that the increase of cancer among men is more rapid than that among women and that during the last decades observed there was a slight fall in the cancer mortality of English women between 35 and 45 years of age.

This more rapid increase at the higher ages is quite probably connected with improved diagnosis and certification. Ignorant or careless certification is likely to affect more seriously deaths not only of the weaker economic classes, like the Negroes, but also of the less economically valuable age periods, like those near the end of a normal lifetime. Hence an improvement in cancer diagnosis and certification would affect especially the increase in the reported death rate at later ages.

8. Is the increase in mortality from such forms of cancer as are difficult to diagnose more rapid than the increase in the forms easy to diagnose?

Digitized by Google

The classification of deaths from cancer made by Messrs. King and Newsholme into accessible and inaccessible cases suffers from one questionable assumption. It tacitly assumes that such a dual classification is possible. In fact cases of cancer fall into a series with those easy of diagnosis at one end and those very difficult of diagnosis at the other. One important element, perhaps the most important element, in determining the position of any specific case on that scale is the part of the body involved. A recent bulletin of the federal Census Bureau* affords an objective test of this element by indicating what proportion of cases of each location were diagnosed by the reporting physician with reasonable certainty. It furnishes what students of the statistics of cancer have lacked hitherto, an objective basis for classifying deaths from this cause according to the certainty of the diagnosis and the part of the body first attacked. This classification has been derived from the physicians who reported 52,420 deaths from this disease in 1914. Their statements showing whether the diagnosis in each case rested on clinical evidence alone or had been confirmed by operation or autopsy enabled the Washington office to divide these 52,420 deaths into those in which the diagnosis was reasonably certain and those in which it was uncertain. From these figures the following results have been derived:

* *Mortality from Cancer and other Malignant Tumors in the Registration Area of the United States, 1914.*

Digitized by Google

DEATHS FROM CANCER CLASSIFIED ACCORDING TO PART OF BODY AFFECTED, SHOWING PER CENT. WITH DIAGNOSIS UNCERTAIN, REGISTRATION AREA OF UNITED STATES, 1914.

Part of Body Affected.	Number of Deaths Excluding Those with Diagnosis Unknown.	Cases where Diagnosis was Uncertain.	
		Number.	Per Cent.
Breast	5,421	1	0
Skin	1,957	0	0
Jaw	851	1	0
Tongue	614	0	0
Lip	376	0	0
Mouth	230	0	0
Vagina and vulva	184	0	0
Testes	121	0	0
Uterus	7,468	8	0.1
Larynx	341	1	0.3
Rectum	2,164	14	0.6
Ovary and fallopian tube	366	56	15.3
Pancreas	567	219	38.6
Kidneys and suprarenals	434	171	39.4
Brain	98	39	39.8
Bladder	764	305	39.9
Peritoneum and mesentery	387	159	41.1
Bones (except jaw)	359	156	43.5
Prostate	625	314	50.2
Intestines	3,121	1,632	52.3
Lung and pleura	278	157	56.5
Pharynx	42	24	57.1
Esophagus	484	285	58.9
Liver and gall bladder	4,777	3,421	71.6
Stomach	9,681	6,973	72.0

These results furnish a test of the agreement between the classification by the American Census Bureau into cases of certain and uncertain diagnosis and the classification by Messrs. King and Newsholme into accessible and inaccessible cases. The English authorities included under accessible only cancers of the tongue, mamma, uterus, and vagina. The foregoing figures confirm this classification in every respect, but suggest that to these might well have been added cancers of other parts of the mouth than the tongue, of the skin, of the larynx, of the testes, and of the rectum. But cancers of the skin, the mouth, and the testes were not given separately in the Frankfort printed sources which they used, and they point out that, in their figures, cancers of the larynx and the rectum were so few or increased so little that a change in the classification of these would not affect the argument.

We may conclude that the dual classification of Messrs. King and Newsholme is justified by these results. A sharp line separates the accessible cases from the inaccessible and to

Digitized by Google

the former class, in addition to those placed there by Messrs. King and Newsholme, belong cancers of the mouth, larynx, skin, rectum, and testes. But the cases of inaccessible cancer may be ranged roughly in a series according to the reported uncertainty of the diagnosis from cancer of the ovary and fallopian tube at one extreme to cancer of the stomach, liver, and gall bladder at the other. If the increase of cancer is largely due to improved diagnosis and the classification of cases according to the part of the body affected is correct and complete, then the rapidity of increase in the reported rate for inaccessible cancer of any particular internal organ and the period through which it persisted would probably vary with the difficulty of diagnosis. It may be wise in future to attempt an improvement of the method introduced by Messrs. King and Newsholme in conformity with this suggestion, but for the present the American material is entirely inadequate for such a study and that from other countries cannot be easily found or interpreted with confidence.

9. Is the rate of cancer mortality lower and its increase faster among men than among women?

The evidence upon which an affirmative answer to both branches of this question is based has already been reviewed in the examination of that part of Messrs. King and Newsholme's argument (pp. 705, ff.).

10. Does the increase of cancer mortality depend upon the initial rate, being high when that rate is small and diminishing or even disappearing when it approaches or reaches a certain high rate?

If the assumption which we are trying to test is correct, if the true death rate from cancer varies only with age and all other variations in the reported rates depend upon the accuracy of those rates and so upon the accuracy of diagnosis, then the increase would be likely to depend upon and vary with the initial rate. So long as the reported rate remained far below the true rate, the increase might be constant or might vary irregularly; but when the reported rate drew near the true rate and the opportunities for further improvements of diagnosis began to disappear, the increase would inevitably slacken. On the other hand, if the spread of the disease were actual, there would be no apparent reason why the disease should

Digitized by Google

grow only towards or to some fixed limit. With the large amount of material recently brought together by the various international compilations, it has become possible to test this inference from our assumption on a scale hitherto quite out of the question.

We may test the theory first in its application to the 22 countries for which the cancer death rates have been computed from data found in *March*.* From these rates quinquennial averages have been computed and the average increase in these per cents. has been determined, the results being given in the following table:

Range of Quinquennial Rates.	Number of Rates Involved.	Average Rate of Increase in Next Five Years.	
		Amount.	Per Cent.
0–29	10	4.7	21.7
30–39	19	6.5	18.9
40–49	20	6.7	15.1
50–59	17	8.6	15.9
60–69	15	8.3	12.8
70–79	6	7.5	9.8
80–89	5	10.7	12.7
90–99	8	5.3	5.5
100+	5	4.0	3.5

* The data speak for the following countries and dates, a total of 694 years.

Country.	Period.	Number of Years.
England and Wales	1858–1910	53
Scotland	1855–1910	56
Ireland	1864–1910	47
Norway	1870–1910	41
Austria	1873–1910	38
Switzerland	1877–1910	34
German Empire	1892–1910	19
Prussia	1875–1910	36
Bavaria	1888–1910	23
Saxony	1892–1910	19
Würtemberg	1892–1910	19
Holland	1875–1910	36
Spain	1900–1910	11
Italy	1887–1910	24
Serbia	1892–1910	19
New South Wales	1875–1910	36
Victoria	1864–1910	47
Queensland	1879–1910	32
South Australia	1879–1910	32
West Australia	1891–1910	20
Tasmania	1891–1910	20
New Zealand	1879–1910	32

The cancer death rates for each country in each year will be found in Appendix, Table XX. In some instances they will be seen to differ slightly from the rates given in *Hoffman*. As the *March* volumes received the coöperation and aid of the statistical offices of the various countries and I have found no reason to question their accuracy, I must ascribe these differences to differences in classification. I believe the early and later figures for each country in *March* are comparable and that is the important matter.

Digitized by Google

With two exceptions the per cent. of increase in the cancer mortality fell as the initial rate rose and after the initial rate reached 80 the average amount of increase also fell rapidly as the initial rate rose.

A similar line of argument has been applied to the larger and more trustworthy body of material in *Falkenburg* and *Hoffman* for cities of at least 100,000 population. Each 5 year rate for each city has been obtained, these rates have been arranged in order of size, the increase (or decrease) between each 5 year rate and that for the following 5 years obtained and from these figures the average increase for each group within certain limits has been found. The results have been brought together in the following summary:

Range of Quinquennial Rates.	Number of Rates Involved.	Average Rate of Increase in Next Five Years.	
		Amount.	Per Cent.
0- 9	11	1.0	19.1
10-19	6	8.4	59.4
20-29	11	8.2	31.7
30-39	32	6.3	17.9
40-49	63	7.2	16.0
50-59	76	7.5	13.6
60-69	83	7.1	11.1
70-79	74	9.6	12.9
80-89	44	6.4	7.5
90-99	46	4.1	4.3
100+	85	4.9	4.2

Here, too, the average per cent. of increase falls almost from the start as the initial rate rises and the amount of increase also falls rapidly after a certain rate, in this instance 70 per 100,000, is reached.

To supplement this information for the leading countries and cities of the world, I have employed the returns of the annual volumes of American *Mortality Statistics* for the 15 year period, 1900–1914. They embrace 49 American cities, each having at least 100,000 inhabitants in 1910, while the preceding tabulation included less than half as many American cities. In this case, instead of computing quinquennial averages, the rate in 1900 has been compared with that in 1914.

Digitized by Google

Range of Initial Rates.	Number of Rates Involved.	Average Rate of Increase in Next Fourteen Years.	
		Amount.	Per Cent.
30–39	3	42.0	116.0
40–49	5	31.3	68.7
50–59	10	31.6	58.6
60–69	18	28.4	44.6
70–79	6	21.7	29.1
80+	7	17.7	20.4

Here, too, it appears that the average per cent. of increase slackens regularly and rapidly as the initial amount rises. The average amount of increase also falls regularly but the change does not appear until a certain rate, in this case 50 per 100,000, is reached.

One more effort to widen the inductive evidence has been made. In *Mortality Statistics, 1909*, the death rates from cancer in each year of the preceding decade is reported for 142 cities which in 1910 had between 20,000 and 100,000 inhabitants. These cities are so small that it is unsafe to take the death rate of a single year as a standard. To avoid the difficulty, I have compared their average cancer death rate 1900–1904 and 1905–1909 city by city. As a shorter period of time is used, the results are correspondingly less important.

Range of Quinquennial Rates.	Number of Rates Involved.	Average Rate of Increase in Next Five Years.	
		Amount.	Per Cent.
0–39	9	12	35
40–59	54	11	22
60–79	50	9	15
80+	29	10	12

The preceding table shows that these smaller cities in general conform to the rule already illustrated in the rates for the foreign countries and larger foreign and American cities. In the death rate of the cities with a low initial rate the absolute increase somewhat exceeds that in cities with a high initial rate and the percentage of increase, which in my judgment is

Digitized by Google

the fairer test, is much more rapid. The results from an extensive study of all these compilations of cancer statistics corroborate each other and constitute, I believe, strong evidence in support of the theory that the apparent increase in cancer mortality is not real.

Arguments from analogy are seldom of value in matters of this sort, but one analogy, that of appendicitis, seems to deserve mention. The opinions of physicians and surgeons regarding the actual increase of this disease diverge as widely as they do regarding the increase of cancer. Yet a confident belief in the increase of appendicitis is probably less general than the same belief regarding cancer, a larger proportion of medical men being ready to ascribe the increase of appendicitis to better diagnosis. Indeed many physicians are surprised to learn that in that part of the United States for which the facts are known the deaths reported from appendicitis are multiplying as fast as deaths from cancer. The following figures show the rates and the increase of each disease in the registration states of 1900 for each year between that date and the present.*

Year.	Death Rate per 100,000 Population.		Ratio of Rate to that of 1900 = 100.	
	Cancer.	Appendicitis.	Cancer.	Appendicitis.
1900	63.9	8.8	100	100
1901	65.8	8.8	103	100
1902	65.6	8.8	103	100
1903	69.0	9.6	108	109
1904	70.4	10.3	110	117
1905	72.4	10.2	113	116
1906	73.1	10.3	114	117
1907	75.7	10.2	118	116
1908	76.8	10.3	120	117
1909	79.4	10.8	124	123
1910	83.0	11.1	130	126
1911	83.9	11.3	131	128
1912	85.9	11.0	134	125
1913	88.7	11.6	139	132
1914	89.1	12.2	139	139
1915	91.8	12.3	144	140

A similar but more rapid spread of appendicitis in England is indicated by the reports. The mortality from it before 1901 is unknown. "The recorded death rate has increased

* For complete figures see Appendix, Table XXXIII.

Digitized by Google

from 38 per 1,000,000 in 1901 to 66 in 1910."* How much of this increase in either country is due to the better diagnosis of a disease which always attacks an inaccessible organ, I see no present means of determining.

SUMMARY OF RESULTS

The reported mortality from cancer is increasing in almost every part of the world, but the real mortality, if increasing at all, is certainly not increasing with equal rapidity.

To determine whether there is a real increase or to justify a denial of it, such parts of the reported increase as are not real must be measured and substracted from the total, in order to ascertain whether there is a residuum not thus to be explained away. For dealing with this problem, corrected, or standardized, death rates are likely to be misleading rather than helpful because the amount and rate of increase shown by them depend largely upon the standard used. This conclusion opens a way for using the data in the international compilations recently published, none of which are in the shape of corrected rates or can readily be converted into them.

A substitute for standardized rates as a way of measuring how much of the increase in the reported death rates from cancer is due to changes in the sex and age composition of a population during any period studied has been applied to 5 cases with the result that between one sixteenth and fourteen sixteenths and, on the average, one third, of the increase was found to be so accounted for. Evidently the part varies widely, but to assume that not less than one tenth of the reported increase of cancer mortality is due to changes in sex and age composition seems a conservative inference.

The remainder, perhaps nine tenths of the increase, must be explained by improved diagnosis, or accepted as actual.

Cancer mortality is lower but increasing faster among men than among women. Both differences may be due to the fact that among men the organs chiefly attacked by cancer are inaccessible while among women they are accessible. Hence the scope for improving the diagnosis of cancer among men was and is much wider.

* England and Wales, *73rd Annual Report of the Registrar General* (1910), p. lxxii.

Digitized by Google

The argument that the Frankfort corrected rates, 1860–1890, indicate that the increase was confined to inaccessible cancer and, therefore, was probably due entirely to improved diagnosis is strengthened by compiling the Frankfort data, 1890–1913, from official manuscript sources, nearly doubling the period observed and trebling the material.

When the death return comes from anyone but a physician, the cause of death is seldom reported as cancer. Hence there is a uniform and close connection between the proportion of deaths certified by physicians and the proportion ascribed to cancer.

The reported cancer mortality is low in those Austrian provinces where medical certificates are few and higher where they are many. A similar relation is suggested in Europe when the countries are arranged in order of cancer mortality, Switzerland having the highest rate and Serbia the lowest. This is of the first importance for the present problem, because in many countries the proportion of deaths certified by physicians is rapidly rising. Even when this is not the case, the proportion of ignorant and incompetent physicians is usually decreasing, a change which would tend to the same result.

At ages when the cancer mortality is highest the proportion of deaths certified by physicians to all deaths is much lower than in early middle life. Hence at those high ages there has been more room for a change from certification by a layman to certification by a physician and probably such changes have been made on a large scale.

The increase in cancer mortality is greatest at the upper ages, when the economic value of a life is least and diagnosis a few decades ago was probably most faulty.

The facilities for diagnosing cancer by microscopical examination of the diseased tissues, etc., are probably better in cities than in the country and better in hospitals than in general practice. Hence the disproportionate increase of urban population, the multiplication of hospital beds and, in many European countries, the increase of autopsies make for greater accuracy in the diagnosis of cancer.

The deaths ascribed to ill-defined or unknown causes in the American registration states of 1900 decreased between 1900

Digitized by Google

and 1915 to less than one tenth of the initial number and the death rate still faster. If these deaths no longer ascribed to ill-defined or unknown causes were distributed over the other causes in a random way, the transfer to cancer accounted for about one ninth of the increase in the reported mortality from that disease. This is probably an understatement of the proportion to be so explained. Similar changes have occurred in many other countries.

The deaths ascribed to tumor in the American registration states of 1900 decreased between 1900 and 1915 to a little over one fourth and the rate to one fourth of the initial figure. Where deaths were reported in 1900 only by enumerators, and so negligently, the proportion ascribed to tumor to all ascribed to tumor or cancer was 7 times as great as within the registration area in the same year. The change from tumor to cancer in the registration states between 1900 and 1914 and in Massachusetts between 1855 and 1912 seems to account for one twentieth of the increase in the cancer death rate. English evidence seems to corroborate the American.

The deaths ascribed to old age in the American registration states of 1900 decreased between 1900 and 1915 to a little more than two fifths, and the rate to less than one third, of the initial figure. The transfers from old age to cancer probably account for at least one eighth of the increase of cancer mortality in those states, 1900–1915.

The measurable influences swelling the increase in the reported mortality from cancer above the true mortality, namely, the increasing proportion of elderly persons in the population, the transfers from ill-defined and unknown causes, from tumor and from old age to cancer seem to account in the American registration states for fully one half of the reported increase, leaving another half to be otherwise explained or accepted as real.

Arguments contributing to explain away this remaining half either cannot yet be stated in the quantitative form, or, if so, cannot be combined with the preceding by any process of addition.

If it should prove to be a general rule, as it apparently is true in Frankfort, that the increase is confined to mortality

Digitized by Google

from inaccessible cancer, that would go far towards establishing the view that the whole increase was only apparent. But notwithstanding fragments of evidence from other districts, this form of argument is still restricted in the main to the Frankfort material.

In cities the cancer mortality is usually higher but increasing more slowly than in country districts and that notwithstanding the larger proportion of aged persons with high cancer mortality in the country. This may be connected with the larger proportion, greater accessibility and expertness of city physicians,—factors which have brought to the city records a large number of cancer deaths which would have been reported otherwise in the country. The difference between the two regions is decreasing and will perhaps continue to decrease, as the improvement in city records slackens as they approach perfection and the improvement in the country records accelerates for a time.

Cancer mortality among Negroes in the United States is lower but increasing faster than among whites. It seems likely that this is due to marked but diminishing differences in correctness of diagnosis.

The increase in cancer mortality varies roughly with the size of the initial rate, tending, after a certain moderately high rate is reached, to decrease as the initial rate rises.

In England and the United States the increase in cancer mortality is parallel with the increase in mortality from appendicitis and both may be due entirely, as they certainly are in large degree, to the improvement of diagnosis.

The cumulative evidence that improvements in diagnosis and changes in age composition explain away more than half and perhaps all of the apparent increase in cancer mortality rebuts the presumption raised by the figures and makes it probable, although far from certain, that cancer mortality is not increasing.

Digitized by Google

APPENDIX

TABLE I (a)

DEATH RATE FROM CANCER CLASSIFIED BY SEX FOR CONNECTICUT, MAINE, MASSACHUSETTS, NEW HAMPSHIRE AND RHODE ISLAND, BY FIVE YEAR PERIODS, 1896–1910.

Period.	Population.	Deaths from Cancer.	Death Rate per 100,000 Population.
	MALE.		
1896–1900	12,462,345	5,755	46.2
1901–1905	13,539,529	7,253	53.6
1906–1910	14,863,173	9,136	61.5
	FEMALE.		
1896–1900	12,839,093	11,592	90.3
1901–1905	13,891,724	14,016	100.9
1906–1910	15,102,303	16,741	110.9

(a) Data from *Hoffman*, pp. 456–469.

TABLE II (a)

DEATH RATE FROM CANCER CLASSIFIED BY SEX, REGISTRATION STATES AS CONSTITUTED IN 1900, (b) FOR YEARS 1900–1915.

Year.	Population.	Deaths from Cancer.	Death Rate per 100,000 Population.
	MALE.		
1900	10,004,754	4,703	47.0
1901	10,222,010	4,982	48.7
1902	10,439,266	4,950	47.4
1903	10,656,522	5,393	50.6
1904	10,873,778	5,631	51.8
1905	11,091,034	5,881	53.0
1906	11,308,290	6,136	54.3
1907	11,525,546	6,454	56.0
1908	11,742,802	6,606	56.3
1909	11,960,058	7,115	59.5
1910	12,177,315	7,626	62.6
1911	12,394,571	7,962	64.2
1912	12,611,827	8,101	64.2
1913	12,829,083	8,617	67.2
1914	13,046,339	8,915	68.3
1915	13,263,595	9,589	72.3

Digitized by Google

FEMALE.

1900	9,990,458	8,066	80.7
1901	10,186,856	8,456	83.0
1902	10,383,254	8,703	83.8
1903	10,579,652	9,257	87.5
1904	10,776,051	9,616	89.2
1905	10,972,450	10,102	92.1
1906	11,168,849	10,290	92.1
1907	11,365,248	10,870	95.6
1908	11,561,647	11,290	97.7
1909	11,758,046	11,715	99.6
1910	11,954,444	12,398	103.7
1911	12,150,842	12,634	104.0
1912	12,347,240	13,350	108.1
1913	12,543,638	13,880	110.7
1914	12,740,037	14,052	110.3
1915	12,936,437	14,472	111.9

(a) Figures from a manuscript table kindly furnished the writer by the Bureau of the Census.
(b) Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, District of Columbia, Indiana, and Michigan.

TABLE III. (a)

DEATH RATE FROM CANCER CLASSIFIED BY SEX, FOR FOURTEEN AMERICAN CITIES, (b) BY FIVE YEAR PERIODS, 1891–1910.

Period.	Population.	Deaths from Cancer.	Death Rate per 100,000 Population.
	MALE.		
1891–1895	12,805,336	5,059	39.4
1896–1900	16,461,766	7,543	45.8
1901–1905	20,382,334	10,631	52.1
1906–1910	23,045,672	14,245	61.8
	FEMALE.		
1891–1895	13,255,922	9,535	71.9
1896–1900	16,962,809	13,173	77.7
1901–1905	20,851,575	17,885	85.8
1906–1910	23,409,743	22,345	95.5

(a) Data from *Hoffman*, pp. 475, ff. The figures for New York City speak for Greater New York since 1898 and for Manhattan and Bronx Boroughs before that date. The cancer mortality of Brooklyn by sex 1891–1898 is unknown.
(b) Augusta, Ga., Boston, Cincinnati, Cleveland, Hartford, Nashville, New Haven, New York City, Philadelphia, Providence, Richmond, Springfield, Mass., St. Louis, Washington.

Digitized by Google

TABLE IV. (a)

DEATH RATE FROM CANCER CLASSIFIED BY SEX, FOR ENGLAND AND WALES, IRELAND, NORWAY, BAVARIA, ITALY, NEW SOUTH WALES, VICTORIA, SOUTH AUSTRALIA, QUEENSLAND, TASMANIA, NEW ZEALAND, BERMUDAS, AND JAMAICA, BY FIVE YEAR PERIODS, 1896–1910.

Period.	Population.	Deaths from Cancer.	Death Rate per 100,000 Population.
	MALE.		
1896–1900	199,454,004	113,988	57.2
1901–1905	207,654,862	132,566	63.8
1906–1910	215,633,636	155,607	72.2
	FEMALE.		
1896–1900	205,838,996	158,280	76.9
1901–1905	214,802,328	178,108	82.3
1906–1910	224,589,489	200,667	89.4

(a) Data from *Hoffman*, pp. 596, ff.

TABLE V. (a)

DEATH RATE FROM CANCER CLASSIFIED BY SEX, FOR CERTAIN FOREIGN CITIES, (b) BY FIVE YEAR PERIODS, 1896–1910.

Period.	Population.	Deaths from Cancer.	Death Rate per 100,000 Population.
	MALE.		
1896–1900	34,316,038	27,581	80.4
1901–1905	36,819,660	33,243	90.3
1906–1910	39,343,588	39,118	99.4
	FEMALE.		
1896–1900	37,160,633	39,996	107.6
1901–1905	40,045,459	46,070	115.0
1906–1910	42,691,732	51,277	120.1

(a) Data from *Hoffman*, pp. 605, ff.
(b) London, Sheffield, Liverpool, Manchester, Cities of Denmark, Bremen, Berlin, Frankfort, Munich, Paris, Budapest, Sydney, Toronto, Buenos Aires.

Digitized by Google

TABLE VI. (a)

DEATHS FROM CANCER IN FRANKFORT-ON-MAIN AT ALL AGES, CLASSIFIED BY PART OF BODY AFFECTED.

Part of Body Affected.	1860–1866		1867–1873		1874–1880		1881–1887		1888–1893(b)		1894–1899(c)		1900–1904		1905–1909		1910–1913		Total.		
	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Both.
Position undefined	33	29	18	30	42	20	58	55	31	32	26	29	26	38	42	74	42	63	318	370	688
Nervous system	5	3	8	5	4	1	10	10	2	4	3	3	4	6		1	2	2	38	35	73
Heart		1													3				3	1	4
Respiratory organs	2	3	2	1	8	4	8	3	8	12	20	11	29	13	39	36	31	23	147	106	253
Tongue	4		4		1	4	7	1	10	1	10	2	8	1	12		13	4	69	13	82
Œsophagus and pharynx	6	2	5	3	15	4	30	11	41	4	62	10	69	17	87	15	51	16	366	82	448
Stomach	36	55	46	58	80	73	104	121	110	108	136	147	212	196	227	263	261	248	1212	1269	2481
Intestines	19	17	18	29	38	37	22	57	54	45	73	48	91	93	135	136	137	93	587	555	1142
Intra and retro-peritoneal	5	5	6	16	9	18	9	20	11	22	12	36	8	31	11	31	6	25	77	204	281
Liver	23	27	36	40	51	57	67	72	33	73	50	106	64	108	88	127	68	128	480	738	1218
Pancreas and spleen		1	1	1	4	1	1	4	1	4	5	3	9	11	8	9	5	4	34	38	72
Kidneys	1	1	4	2	4	2	7	3	7	1	9	3	6	3	10	5	5	8	53	28	81
Prostate, urinary bladder and penis	7	1	10	4	7	1	14	5	9	5	24	17	33	9	30	11	34	20	168	73	241
Uterus		69		96		117		159		158		148		182		206		161		1296	1296
Ovaries		9		9		17		20		30		14		8		14		13		134	134
Mamma		43		42		60		41	1	63		55		72		83	1	101	2	560	562
Vagina		1		2		5		5		5		4		2		3		20		47	47
Bone	2		2		5	1	5	1	4	4	13	11	13	13	16	23	12	14	72	67	139
Total	143	267	160	338	268	422	342	588	322	571	443	647	572	803	708	1037	668	943	3626	5616	9242

(a) Figures from 1860 to 1887, inclusive, derived from *King and Newsholme*, Table XIII. Figures for later years derived from manuscript tables from the Frankfort Statistical Office.
(b) Omitting the year 1890, for which the figures were not available.
(c) Omitting the year 1896, for the above reason.

Digitized by Google

TABLE VII. (a)

DEATHS FROM CANCER IN FRANKFORT-ON-MAIN OVER TWENTY YEARS OF AGE.

Age.	1860–1866		1867–1873		1874–1880		1881–1887		1888–1893 (b)	
	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.
20–29	3	5	5	3	7	5	6	15	1	9
30–39	16	13	17	32	19	40	34	50	15	53
40–49	18	51	20	65	47	92	57	123	66	101
50–59	50	86	40	95	60	112	94	143	106	166
60–69	36	70	47	80	79	99	88	151	80	153
70–79	19	35	26	54	43	67	46	75	47	66
80+	4	4	3	5	5	5	5	20	6	17
Total.....	140	264	158	334	260	420	330	577	321	565

Age.	1894–1899 (c)		1900–1904		1905–1909		1910–1913		Total.		
	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Both.
20–29	8	5	9	13	12	12	11	9	62	76	138
30–39	26	48	41	59	25	55	41	70	228	420	648
40–49	68	130	73	130	88	179	99	160	536	1,031	1,567
50–59	138	188	148	238	208	284	145	246	989	1,558	2,547
60–69	123	175	189	215	255	298	238	260	1,135	1,501	2,636
70–79	62	84	90	121	95	166	118	158	546	826	1,372
80+	10	14	17	25	15	34	7	32	72	156	228
Total	435	644	567	801	698	1,028	659	935	3,568	5,568	9,136

(a) Figures from 1860 to 1887, inclusive, derived from *King and Newsholme*, Table XIV. Figures for later years derived from manuscript tables from the Frankfort Statistical Office.
(b) Omitting the year 1890, for which the figures were not available.
(c) Omitting the year 1896, for which the figures were not available.

TABLE VIII. (a)

DEATHS FROM CANCER IN FRANKFORT-ON-MAIN OVER TWENTY YEARS OF AGE CLASSIFIED INTO ACCESSIBLE CANCER, INACCESSIBLE CANCER AND CANCER IN UNDEFINED POSITION.

Age.	1860–1866		1867–1873		1874–1880		1881–1887		1888–1893 (b)		1894–1899 (c)		1900–1904		1905–1909		1910–1913		Total.		
	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Males.	Females.	Both.
Accessible Cancer.																					
20–29		1				4		1			1	5		4		4	1	19	20		
30–39		6	1	19		21	1	26	1	30		20		29		27		30	3	208	211
40–49		36	2	28		53	4	63	7	55	4	53	1	59	5	70	5	78	28	495	523
50–59	6	40		51		51	1	62	4	71	7	71	5	87	6	96	7	69	36	598	634
60–69	3	25	6	30		46		34	2	53	3	48	2	58	5	72	6	71	27	437	464
70–79		7		14	1	14	1	13	1	15		15	2	22	2	27	4	33	11	160	171
80+	1			1		1		4		4		4		6		9	2	10	3	39	42
Total...	10	115	9	143	1	186	7	206	15	229	14	211	11	266	18	305	24	296	109	1956	2065

Digitized by Google

INACCESSIBLE CANCER.

20–29	1	4	3	3	6	5	1	10		8	6	5	8	7	11	7	10	4	46	53	99
30–39	4	6	13	11	17	15	21	21	11	20	22	27	39	26	23	25	37	32	187	183	370
40–49	14	9	15	33	40	36	44	46	51	41	63	72	67	66	77	96	89	70	460	469	929
50–59	34	36	35	34	50	54	80	64	93	91	122	108	141	146	191	167	134	164	880	864	1744
60–69	27	39	38	44	71	50	78	104	75	94	118	117	182	151	242	212	224	175	1055	986	2041
70–79	16	23	25	36	32	50	42	59	43	45	59	67	84	93	89	133	107	120	497	626	1123
80+	1	4	3	3	4	4	3	14	6	11	10	10	15	17	15	24	4	22	61	109	170
Total..	97	121	132	164	220	214	269	318	279	310	400	406	536	506	648	664	605	587	3186	3290	6476

CANCER, POSITION UNDEFINED.

20–29	2		2		1		5	1	1		2			1	1	1	1	1	15	4	19
30–39	6	1	3	2	2	4	12	3	3	3	4	1	2	4	2	3	4	8	38	29	67
40–49	4	6	3	4	7	3	9	14	8	5	1	5	5	5	6	13	5	12	48	67	115
50–59	10	10	5	10	10	7	13	17	9	4	9	9	2	5	11	21	4	13	73	96	169
60–69	6	6	3	6	8	3	10	13	3	6	2	10	5	6	8	14	8	14	53	78	131
70–79	3	5	1	4	10	3	3	3	3	6	3	2	4	6	4	6	7	5	38	40	78
80+	2			1	1		2	2		2			2	2		1	1		8	8	16
Total...	33	28	17	27	39	20	54	53	27	26	21	27	20	29	32	59	30	53	273	322	595

(a) Figures from 1860 to 1887, inclusive, derived from *King and Newsholme*, Table XIV continued. Figures for later years derived from manuscript tables from the Frankfort Statistical Office.
(b) Omitting the year 1890, for which the figures were not available.
(c) Omitting the year 1896, for the same reason.

TABLE IX. (a)

POPULATION AT RISK IN FRANKFORT-ON-MAIN.

Age.	1860–1866	1867–1873	1874–1880	1881–1887	1888–1893 (b)	1894–1899 (c)	1900–1904	1905–1909	1910–1913
MALES.									
20–29	90,404	76,174	90,222	96,183	97,888	134,459	173,683	192,245	169,211
30–39	42,294	49,166	69,410	87,654	70,964	93,702	124,430	154,841	158,092
40–49	29,201	30,964	40,993	57,471	51,794	65,878	77,958	93,107	97,586
50–59	18,924	30,036	23,860	30,224	29,154	41,070	51,297	57,152	53,499
60–69	10,850	11,556	12,918	16,158	13,586	18,124	24,008	29,540	29,912
70–79	3,700	4,228	5,116	6,292	5,109	6,506	7,898	9,240	9,984
80+	617	750	919	1,137	842	1,091	1,319	1,454	1,381
Total...	195,990	192,874	243,438	295,119	269,337	360,830	460,593	537,579	519,665
FEMALES.									
20–29	74,536	85,365	108,009	133,871	119,057	157,951	193,103	213,162	192,782
30–39	40,880	50,361	70,137	91,459	76,921	99,789	126,516	154,969	156,157
40–49	27,260	31,019	41,705	60,116	54,239	69,218	82,207	97,156	98,545
50–59	19,645	21,320	26,813	34,575	33,308	46,071	56,964	64,094	61,067
60–69	12,172	13,881	16,524	21,439	18,143	24,286	31,934	38,166	37,998
70–79	4,272	5,288	7,008	8,957	7,774	9,975	11,951	14,077	14,508
80+	1,110	944	1,077	1,904	1,632	1,879	2,296	2,630	2,490
Total..	179,875	208,178	271,273	352,321	311,074	409,169	504,971	584,254	563,547

(a) Figures from 1860 to 1887, inclusive, derived from *King and Newsholme*, Table XV. Figures for later years derived from German census report.
(b) Omitting the year 1890, for which the figures were not available.
(c) Omitting the year 1896, for which the figures were not available.

Digitized by Google

TABLE X. (a)

FRANKFORT. ANNUAL DEATHS FROM CANCER IN 1,000,000 LIVING, AT LEAST TWENTY-FIVE YEARS OF AGE. POPULATION DISTRIBUTED IN AGE GROUPS ACCORDING TO ENGLISH LIFE TABLE NO. 3—PERSONS.

Period.	Accessible.	Inaccessible.	Position Undefined.	Total.
		MALES.		
1860–1866	126	1,118	359	1,603
1867–1873	88	1,421	127	1,646
1874–1880	14	1,913	363	2,290
1881–1887	35	1,865	305	2,205
1888–1893 (b)	91	2,289	171	2,551
1894–1899 (c)	66	2,489	104	2,659
1900–1904	51	2,742	118	2,911
1905–1909	68	2,791	119	2,981
1910–1913	119	2,522	130	2,771
		FEMALES.		
1860–1866	1,081	1,323	263	2,697
1867–1873	1,214	1,540	254	3,008
1874–1880	1,220	1,588	131	2,939
1881–1887	981	1,820	272	3,073
1888–1893 (b)	1,275	1,959	174	3,408
1894–1899 (c)	903	1,936	122	2,961
1900–1904	914	2,013	115	3,042
1905–1909	925	2,311	180	3,416
1910–1913	913	2,057	152	3,122

(a) Figures from 1860 to 1887, inclusive, derived from *King and Newsholme*, Table XVI. Figures for later years derived from manuscript tables from the Frankfort Statistical Office.
(b) Omitting the year 1890, for which the figures were not available.
(c) Omitting the year 1896, for which the figures were not available.

TABLE XI.

STANDARD MILLION, ENGLAND AND WALES, AS SHOWN BY ENGLISH LIFE TABLE NO. 3 AND BY CENSUS OF 1881.

Age Period.	English Life Table No. 3.	Census of 1881.
	MALES.	
Under 5	49,630	67,669
5–9	42,710	60,389
10–14	41,210	53,985
15–19	40,150	48,828
20–24	38,700	42,825
25–29	36,980	37,238
30–34	35,190	32,890
35–39	33,300	29,095
40–44	31,250	25,493
45–49	28,990	21,640
50–54	26,430	18,140
55–59	23,470	15,550
60–64	20,070	12,268
65–69	16,110	8,946
70–74	11,670	6,064
75–79	7,230	3,541
80–84	3,590	1,503
85–89	1,320	448
90–94	330	101
95+	50	15
	488,380	486,628

Digitized by Google

FEMALES.

Under 5	51,000	67,882
5–9	44,320	60,783
10–14	42,740	53,826
15–19	41,570	49,239
20–24	39,990	46,810
25–29	38,140	40,465
30–34	36,210	35,455
35–39	34,190	31,190
40–44	32,100	27,439
45–49	29,920	23,660
50–54	27,630	20,236
55–59	25,010	17,309
60–64	21,790	13,935
65–69	17,910	10,507
70–74	13,400	7,301
75–79	8,700	4,431
80–84	4,590	2,000
85–89	1,820	670
90–94	500	197
95+	90	37
	511,620	513,372

TABLE XII.

STANDARDIZED CANCER DEATH RATES FOR FRANKFORT, BASED ON ENGLISH LIFE TABLE NO. 3 AND ON ENGLISH CENSUS OF 1881.

Age.	Standard Million.		Cancer Death Rate Frankfort 1860–66. (a)	Computed Cancer Deaths.	
	English Life Table No. 3.	English Census of 1881.		English Life Table No. 3.	English Census of 1881.
			MALES.		
Under 20	173,700	230,871	2.7	5	6
20–29	75,680	80,063	3.5	3	3
30–39	68,490	61,985	20.5	14	13
40–49	60,240	47,133	53.1	32	25
50–59	49,900	33,690	225.1	112	76
60–69	36,180	21,214	278.7	101	59
70–79	18,900	9,605	453.9	86	44
80+	5,290	2,067	514.8	27	11
Total.....	488,380	486,628		380	237
			FEMALES.		
Under 20	179,630	231,730	3.0	5	7
20–29	78,130	87,275	6.0	5	5
30–39	70,400	66,645	27.0	19	18
40–49	62,020	51,099	156.2	97	80
50–59	52,640	37,545	372.4	196	140
60–69	39,700	24,442	477.8	190	117
70–79	22,100	11,732	693.5	153	81
80+	7,000	2,904	332.3	23	10
Total.....	511,620	513,372		688	458
Both sexes...	1,000,000	1,000,000		1,068	695

(a) Computed, using census population of Frankfort, 1864, from *Stat. d. deut. Reichs*, "Beiträge zur Statistik d. St. Frankfurt-am-Main," II (1866), pt. I, Table XIX, and deaths from *King and Newsholme*, p. 240, supplemented as to deaths under 20 from personal investigation.

Digitized by Google

TABLE XIII.

STANDARDIZED CANCER DEATH RATES FOR FRANKFORT, 1888-89, BASED ON ENGLISH LIFE TABLE NO. 3 AND ON ENGLISH CENSUS OF 1881.

Age.	Standard Million.		Cancer Death Rate per 100,000 Population, Frankfort, 1888-89. (a)	Computed Cancer Deaths.	
	English Life Table No. 3	English Census of 1881.		English Life Table No. 3.	English Census of 1881.
Males.					
Under 20	173,700	230,871			
20-29	75,680	80,063	2.7	2	2
30-39	68,490	61,985	14.8	10	9
40-49	60,240	47,133	132.6	80	62
50-59	49,900	33,690	362.0	180	122
60-69	36,180	21,214	466.2	169	99
70-79	18,900	9,605	821.4	155	79
80+	5,290	2,067	625.0	33	13
Total.....	488,380	486,628		629	386
Females.					
Under 20	179,630	231,730	3.0	5	7
20-29	78,130	87,275	6.7	5	6
30-39	70,400	66,645	68.6	48	46
40-49	62,020	51,099	200.0	124	102
50-59	52,640	37,545	493.4	260	185
60-69	39,700	24,442	887.1	352	217
70-79	22,100	11,732	815.8	180	96
80+	7,000	2,904	937.5	66	27
Total.....	511,620	513,372		1,040	686
Both sexes	1,000,000	1,000,000		1,669	1,072

(a) The death rates from cancer 1888-89 were computed by using the population of Frankfort December 1, 1889 (estimated by arithmetical method from population of 1885 and 1890) and the average annual deaths from cancer 1888-89, derived from *King and Newsholme*, Table XIV (this Appendix, Table VII).

TABLE XIV. (a)

DEATH RATE FROM CANCER IN THE REGISTRATION STATES OF 1900 IN THE YEARS 1900-1910.

Year.	Population.	Deaths from Cancer.	Death Rate per 100,000 Population.
1900	19,995,212	12,769	63.9
1901	20,408,866	13,438	65.8
1902	20,822,520	13,653	65.6
1903	21,236,174	14,650	69.0
1904	21,649,829	15,247	70.4
1905	22,063,484	15,983	72.4
1906	22,477,139	16,426	73.1
1907	22,890,794	17,324	75.7
1908	23,304,449	17,896	76.8
1909	23,718,104	18,830	79.4
1910	24,131,759	20,024	83.0

Increase per 100,000 population, 1900-1910.....19.1
Per cent. increase, 1900-1910.....29.9

(a) Data kindly furnished the writer in manuscript by the Bureau of the Census.

Digitized by Google

TABLE XV.

STANDARDIZED CANCER DEATH RATE 1910 IN REGISTRATION STATES AS CONSTITUTED IN 1900. (a)

Age Period.	Males.			Females.		
	Population 1900.	Cancer Death Rate per 100,000 Population 1910.	Estimated Deaths 1910.	Population 1900.	Cancer Death Rate per 100,000 Population 1910.	Estimated Deaths 1910.
Under 5	1,048,490	4.1	43	1,031,197	2.8	29
5–9	1,001,754	1.5	15	988,707	1.2	12
10–14	915,572	1.8	16	909,552	1.4	13
15–19	885,871	2.9	26	926,057	3.5	32
20–24	920,072	4.9	45	993,194	4.1	41
25–34	1,751,535	9.5	166	1,730,952	21.9	379
35–44	1,408,030	33.0	465	1,325,608	88.9	1,178
45–54	966,630	106.7	1,031	928,077	230.7	2,141
55–64	620,054	272.0	1,686	632,720	411.3	2,602
65–74	343,391	493.6	1,695	362,034	616.2	2,231
75+	143,355	693.7	994	162,360	867.8	1,409
Total.....	10,004,754		6,182	9,990,458		10,067

Total population..... 19,995,212
Estimated deaths..... 16,249
Standardised death rate..... 81.3

(a) Population 1900 and cancer death rates 1910 kindly furnished the writer by the Bureau of the Census.

TABLE XVI.

STANDARDIZED CANCER DEATH RATE, ENGLAND AND WALES, 1911.

Age Period.	Males.			Females.		
	Population (a) 1901.	Cancer Death Rate per 100,000 Population (b) 1911.	Estimated Deaths 1911.	Population (a) 1901.	Cancer Death Rate per 100,000 Population (b) 1911.	Estimated Deaths 1911.
Under 5	1,855,361	3.3	61	1,861,347	2.6	48
5–9	1,738,993	2.8	49	1,748,298	1.8	31
10–14	1,670,970	2.2	37	1,670,770	1.3	22
15–19	1,607,522	3.6	58	1,638,621	2.7	44
20–24	1,472,644	6.3	93	1,648,278	4.4	73
25–29	1,328,288	8.5	113	1,496,221	8.0	120
30–34	1,157,666	14.0	162	1,273,665	25.0	318
35–39	1,034,459	27.4	283	1,110,924	56.8	631
40–44	897,484	58.8	528	953,138	113.3	1,080
45–49	759,955	127.2	967	813,233	191.8	1,559
50–54	636,254	224.3	1,427	692,749	264.1	1,829
55–59	497,498	350.0	1,741	555,079	389.4	2,161
60–64	410,447	523.7	2,150	480,226	507.2	2,436
65–69	282,403	695.6	1,964	347,270	622.6	2,162
70–74	195,465	868.4	1,697	250,868	808.1	2,027
75–79	113,096	947.0	1,071	151,384	889.0	1,346
80–84	52,137	851.0	444	76,631	863.8	662
85+	17,971	738.9	133	30,528	808.4	247
Total...	15,728,613		12,978	16,799,230		16,796

Total population..... 32,527,843
Estimated deaths..... 29,774
Standardised death rate..... 91.5

(a) From *Census of England and Wales*, 1901, *Summary Tables*, p. 139.
(b) Computed, using population for 1911 from *Census of England and Wales, 1911*, VII, pp. 1, 2; and deaths for 1911 from *74th Ann. Rep. of Registrar-General* (1911), p. 201.

Digitized by Google

TABLE XVII.

STANDARDIZED CANCER DEATH RATE OF ENGLAND AND WALES, 1891-1900.

Age Period.	Males.			Females.		
	Standard Million 1881-1890.	Cancer Death Rate per 100,000 Population 1891-1900.	Estimated Deaths 1891-1900.	Standard Million 1881-1890.	Cancer Death Rate per 100,000 Population 1891-1900.	Estimated Deaths 1891-1900.
Under 5	64,122	3.3	2	64,557	2.8	2
5-9	59,333	1.8	1	59,673	1.4	1
10-14	54,806	1.9	1	54,765	1.4	1
15-19	49,720	3.2	2	50,287	2.7	1
20-24	42,922	5.1	2	47,564	3.9	2
25-34	71,131	9.9	7	77,499	17.5	14
35-44	55,095	38.4	21	58,944	89.1	53
45-54	40,472	130.0	53	44,478	232.3	103
55-64	27,151	316.0	86	30,893	409.9	127
65-74	15,184	532.5	81	18,326	582.9	107
75+	5,591	582.4	33	7,487	637.7	48
Total.....	485,527		289	514,473		459

Standardised death rate per 100,000..... 74.8

(a) Standard million from *Supp. to 65th Ann. Rep. of Registrar-General* (1891-1900), p. xi; average annual deaths per million, *Ibid.*, p. cxcv.

TABLE XVIII.

STANDARDIZED CANCER DEATH RATE, NEW YORK CITY, 1910.

Age Period.	Males.			Females.		
	Population 1900.	Cancer Death Rate per 100,000 Population (a) 1910.	Estimated Deaths 1910.	Population 1900.	Cancer Death Rate per 100,000 Population (a) 1910.	Estimated Deaths 1910.
Under 5	199,683	7.0	14	197,604	2.8	6
5-9	177,591	0.9	2	177,156	2.3	4
10-14	149,906	1.4	2	151,358	1.4	2
15-19	140,670	4.1	6	162,081	3.3	5
20-24	161,988	9.9	16	192,853	2.8	5
25-34	343,178	12.4	43	338,175	26.4	89
35-44	261,095	44.3	116	232,950	107.9	251
45-54	146,495	172.9	253	141,837	281.4	399
55-64	78,692	398.4	314	82,591	497.9	411
65-74	32,915	733.2	241	39,314	665.0	261
75+	9,799	681.4	67	13,835	884.6	122
Total.....	1,702,012		1,074	1,729,754		1,555

Total population..... 3,431,766
Estimated deaths..... 2,629
Standardised death rate..... 76.6

(a) Death rates kindly furnished in manuscript by Bureau of the Census.

Digitized by Google

TABLE XIX.

STANDARDIZED CANCER DEATH RATE, LONDON, (a) 1911.

Age Period.	Males.			Females.		
	Population 1901. (b)	Cancer Death Rate per 100,000 Population (c) 1911.	Estimated Deaths 1911.	Population 1901. (b)	Cancer Death Rate per 100,000 Population (c) 1911.	Estimated Deaths 1911.
Under 15	675,970	3.5	24	681,904	2.3	16
15–24	426,271	5.5	23	493,478	3.8	19
25–34	371,418	11.8	44	437,763	16.0	70
35–39	150,706	30.4	46	168,882	60.2	102
40–44	129,992	64.0	83	141,486	128.0	181
45–49	107,320	164.9	177	119,010	170.5	203
50–54	87,913	279.5	246	99,284	268.1	266
55–59	65,784	439.3	289	75,821	392.5	298
60–64	52,932	664.3	352	65,153	508.2	331
65–69	33,341	855.0	285	45,125	591.9	267
70–74	21,662	1,130.3	245	32,647	764.7	250
75–79	11,816	1,104.1	130	19,612	942.7	185
80–84	5,218	993.6	52	10,200	963.6	98
85+	1,742	624.2	11	4,091	997.3	41
Total.....	2,142,085		2,007	2,394,456		2,327

Total population	4,536,541
Estimated deaths	4,334
Standardised death rate	95.5

(a) Administrative, or Registration, County.
(b) From *Census of England and Wales, 1911, Summary Tables*, pp. 148–149.
(c) From 74*th Ann. Rep. of Registrar-General* (1901), p. lxxvi.

TABLE XX. (a)

DEATH RATES FROM CANCER IN VARIOUS EUROPEAN COUNTRIES AND THE AUSTRALIAN COLONIES.

Year.	Death Rate from Cancer per 100,000 Population.	Year.	Death Rate from Cancer per 100,000 Population.	Year.	Death Rate from Cancer per 100,000 Population.	Year.	Death Rate from Cancer per 100,000 Population.
			England and Wales. (b)				
1858	32.9	1872	43.1	1886	59.0	1900	83.9
1859	33.7	1873	44.4	1887	61.5	1901	84.3
1860	34.1	1874	46.1	1888	62.2	1902	84.6
1861	36.0	1875	47.1	1889	65.6	1903	87.4
1862	36.1	1876	47.3	1890	67.6	1904	88.2
1863	36.1	1877	48.8	1891	69.2	1905	88.9
1864	38.6	1878	50.3	1892	69.2	1906	92.2
1865	37.2	1879	50.1	1893	71.0	1907	91.5
1866	38.5	1880	51.4	1894	71.2	1908	93.3
1867	39.2	1881	52.0	1895	75.3	1909	96.1
1868	40.3	1882	53.4	1896	76.4	1910	96.7
1869	41.7	1883	54.9	1897	78.5		
1870	42.4	1884	56.5	1898	79.9		
1871	42.4	1885	57.2	1899	82.6		

(a) Computed from figures for population and deaths in *March*, Vols. I and II.
(b) Figures going back to 1838 may be found in the successive *Annual Reports of the Registrar-General.*

Digitized by Google

Scotland.

1855	29.2	1869	43.1	1883	53.6	1897	74.7
1856	30.3	1870	41.1	1884	55.1	1898	79.5
1857	31.9	1871	42.5	1885	56.3	1899	81.3
1858	31.8	1872	45.8	1886	59.5	1900	78.9
1859	33.0	1873	44.7	1887	60.6	1901	81.8
1860	33.2	1874	47.3	1888	62.1	1902	82.3
1861	38.0	1875	47.6	1889	66.5	1903	83.8
1862	37.2	1876	47.7	1890	60.7	1904	85.9
1863	37.2	1877	48.2	1891	67.0	1905	90.0
1864	41.2	1878	50.4	1892	66.6	1906	97.6
1865	39.5	1879	49.9	1893	68.3	1907	97.9
1866	41.3	1880	50.5	1894	70.3	1908	98.5
1867	39.7	1881	51.1	1895	71.1	1909	101.6
1868	40.5	1882	54.5	1896	70.8	1910	102.7

Ireland.

1864	26.6	1876	32.2	1888	41.7	1900	60.8
1865	28.6	1877	35.4	1889	44.9	1901	65.1
1866	30.7	1878	36.2	1890	45.5	1902	64.5
1867	30.2	1879	33.9	1891	46.2	1903	69.0
1868	30.1	1880	34.1	1892	47.9	1904	69.4
1869	30.5	1881	37.1	1893	49.5	1905	74.9
1870	33.1	1882	36.9	1894	51.8	1906	79.2
1871	32.1	1883	39.7	1895	50.4	1907	76.2
1872	32.6	1884	39.1	1896	53.7	1908	75.7
1873	32.7	1885	39.0	1897	58.2	1909	80.0
1874	33.4	1886	41.4	1898	58.8	1910	83.7
1875	33.2	1887	42.6	1899	59.0		

Norway.

1870	27.1	1881	45.7	1891	61.3	1901	95.1
1871	27.3	1882	46.2	1892	63.7	1902	91.5
1872	27.0	1883	49.4	1893	69.7	1903	93.2
1873	34.5	1884	51.6	1894	71.4	1904	96.0
1874	31.9	1885	53.1	1895	71.2	1905	98.4
1875	34.6	1886	50.4	1896	81.1	1906	97.6
1876	34.4	1887	55.6	1897	85.4	1907	100.3
1877	37.1	1888	55.2	1898	84.2	1908	97.2
1878	37.6	1889	56.4	1899	89.1	1909	95.2
1879	41.8	1890	57.0	1900	91.3	1910	92.9
1880	41.4						

Austria.

1873	33.0	1883	44.4	1893	59.0	1902	74.2
1874	33.9	1884	45.7	1894	60.8	1903	73.7
1875	35.2	1885	47.6	1895	63.7	1904	74.9
1876	36.4	1886	48.3	1896	65.8	1905	76.2
1877	37.4	1887	47.8	1897	67.9	1906	77.9
1878	38.4	1888	49.3	1898	69.4	1907	77.3
1879	40.7	1889	51.6	1899	69.9	1908	78.0
1880	40.7	1890	52.6	1900	70.9	1909	78.7
1881	42.2	1891	54.3	1901	72.9	1910	77.9
1882	42.7	1892	57.0				

Switzerland.

1877	78.9	1886	111.5	1895	122.2	1903	129.7
1878	82.4	1887	110.4	1896	120.0	1904	128.6
1879	86.6	1888	113.8	1897	123.4	1905	127.4
1880	90.7	1889	111.7	1898	123.3	1906	129.1
1881	99.3	1890	111.8	1899	122.0	1907	122.5
1882	99.1	1891	114.4	1900	125.4	1908	128.0
1883	101.0	1892	119.8	1901	127.9	1909	126.7
1884	103.9	1893	116.0	1902	125.8	1910	123.5
1885	106.5	1894	116.2				

Digitized by Google

GERMAN EMPIRE. (a)

1892	61.0	1897	69.3	1902	75.0	1907	83.2
1893	63.0	1898	70.4	1903	77.2	1908	83.7
1894	64.5	1899	73.7	1904	79.8	1909	84.5
1895	65.7	1900	72.1	1905	80.9	1910	87.9
1896	68.1	1901	74.7	1906	81.3		

PRUSSIA.

1875	22.9	1884	35.3	1893	51.1	1902	61.9
1876	24.2	1885	35.8	1894	52.8	1903	65.2
1877	26.5	1886	38.3	1895	53.2	1904	68.6
1878	28.2	1887	38.1	1896	54.8	1905	69.4
1879	28.8	1888	40.9	1897	56.0	1906	70.4
1880	29.9	1889	43.5	1898	56.3	1907	73.4
1881	31.1	1890	43.3	1899	59.3	1908	73.6
1882	31.9	1891	44.7	1900	59.6	1909	74.8
1883	33.8	1892	49.6	1901	61.7	1910	78.5

BAVARIA.

1888	79.4	1894	92.6	1900	99.3	1906	110.5
1889	81.6	1895	94.9	1901	102.9	1907	106.8
1890	81.0	1896	95.3	1902	103.4	1908	108.2
1891	88.7	1897	96.9	1903	107.4	1909	109.7
1892	85.0	1898	97.3	1904	110.2	1910	114.2
1893	90.0	1899	101.9	1905	108.9		

SAXONY. (b)

1892	85.7	1897	95.2	1902	97.0	1907	97.3
1893	85.7	1898	91.2	1903	91.0	1908	94.3
1894	89.1	1899	94.8	1904	88.2	1909	93.9
1895	89.5	1900	93.6	1905	91.6	1910	93.9
1896	93.2	1901	97.4	1906	93.8		

WURTEMBERG.

1892	71.4	1897	80.0	1902	97.4	1907	102.8
1893	71.4	1898	95.0	1903	98.4	1908	102.9
1894	70.7	1899	94.0	1904	98.7	1909	105.5
1895	75.2	1900	94.0	1905	104.2	1910	107.3
1896	82.1	1901	94.4	1906	99.3		

HOLLAND.

1875	47.9	1884	61.6	1893	80.8	1902	95.0
1876	48.5	1885	66.0	1894	81.0	1903	98.9
1877	50.2	1886	67.0	1895	85.4	1904	97.9
1878	51.1	1887	65.3	1896	88.5	1905	101.2
1879	52.5	1888	69.5	1897	90.4	1906	100.7
1880	54.4	1889	75.3	1898	93.0	1907	101.8
1881	57.6	1890	73.4	1899	96.0	1908	102.8
1882	58.4	1891	79.4	1900	91.8	1909	102.6
1883	58.0	1892	79.9	1901	93.7	1910	106.4

(a) Figures include about 94 per cent. of the population. For details see notes in *March*, I and II.
(b) For figures running back to 1873 see *Zeitschr. d. k. Sächs. Stat. Landesamtes* 1905, p. 263.

Digitized by Google

SPAIN.

1900	39.3	1903	44.1	1906	47.6	1909	51.0
1901	42.4	1904	46.6	1907	47.5	1910	52.3
1902	43.3	1905	45.8	1908	51.4		

ITALY.

1887	42.7	1893	42.9	1899	51.9	1905	58.3
1888	42.3	1894	44.5	1900	52.2	1906	62.0
1889	43.0	1895	48.2	1901	52.7	1907	61.7
1890	42.7	1896	49.1	1902	53.9	1908	64.5
1891	43.0	1897	50.3	1903	53.8	1909	64.2
1892	42.6	1898	51.1	1904	57.1	1910	65.6

SERBIA.

1895	5.5	1899	9.5	1903	9.1	1907	13.3
1896	6.6	1900	9.5	1904	10.8	1908	12.6
1897	7.0	1901	9.1	1905	10.4	1909	13.1
1898	7.4	1902	9.6	1906	10.6	1910	12.8

NEW SOUTH WALES.

1875	30.9	1884	26.4	1893	40.7	1902	62.3
1876	27.5	1885	28.8	1894	42.1	1903	65.4
1877	26.6	1886	34.4	1895	44.4	1904	66.0
1878	29.0	1887	35.2	1896	48.9	1905	65.3
1879	24.7	1888	39.0	1897	53.3	1906	69.0
1880	32.8	1889	36.9	1898	54.4	1907	70.8
1881	28.2	1890	35.6	1899	57.1	1908	67.5
1882	26.9	1891	45.2	1900	56.5	1909	72.5
1883	25.7	1892	43.3	1901	61.7	1910	71.7

VICTORIA.

1864	21.7	1876	38.0	1888	49.4	1900	68.5
1865	24.7	1877	40.7	1889	57.7	1901	73.4
1866	18.4	1878	38.3	1890	56.0	1902	70.4
1867	18.6	1879	45.9	1891	60.9	1903	76.1
1868	27.6	1880	45.1	1892	58.8	1904	73.9
1869	26.8	1881	40.4	1893	62.6	1905	78.6
1870	30.0	1882	41.9	1894	63.1	1906	75.5
1871	26.2	1883	49.5	1895	64.2	1907	82.1
1872	30.0	1884	47.8	1896	66.7	1908	81.8
1873	32.6	1885	46.5	1897	65.5	1909	82.6
1874	34.5	1886	50.4	1898	73.1	1910	82.5
1875	39.1	1887	51.8	1899	71.0		

QUEENSLAND.

1880	23.0	1888	23.3	1896	39.3	1904	57.2
1881	29.4	1889	32.7	1897	39.1	1905	66.7
1882	22.7	1890	28.2	1898	46.5	1906	54.7
1883	26.1	1891	33.1	1899	47.8	1907	66.4
1884	31.4	1892	31.3	1900	47.7	1908	61.0
1885	16.7	1893	28.3	1901	55.0	1909	60.7
1886	27.6	1894	35.8	1902	55.0	1910	67.3
1887	22.8	1895	41.7	1903	49.1		

Digitized by Google

SOUTH AUSTRALIA.

1880	28.1	1888	37.8	1896	52.5	1904	63.4
1881	32.6	1889	43.1	1897	51.8	1905	69.2
1882	31.0	1890	41.4	1898	51.1	1906	76.9
1883	29.2	1891	49.1	1899	57.0	1907	74.5
1884	35.9	1892	45.5	1900	59.8	1908	73.6
1885	32.8	1893	50.9	1901	60.5	1909	80.0
1886	34.1	1894	47.0	1902	75.0	1910	80.7
1887	36.0	1895	47.4	1903	73.5		

WEST AUSTRALIA.

1895	27.2	1899	35.5	1903	41.6	1907	50.0
1896	25.2	1900	29.7	1904	44.3	1908	54.3
1897	34.0	1901	44.4	1905	50.8	1909	68.0
1898	33.3	1902	41.3	1906	59.2	1910	49.5

TASMANIA.

1895	48.7	1899	53.2	1903	56.2	1907	62.9
1896	60.1	1900	53.8	1904	52.0	1908	67.6
1897	50.0	1901	54.6	1905	54.2	1909	67.4
1898	59.3	1902	61.9	1906	61.9	1910	66.5

NEW ZEALAND.

1880	28.1	1888	43.5	1896	55.1	1904	67.6
1881	27.2	1889	42.5	1897	54.7	1905	65.1
1882	28.9	1890	47.5	1898	64.0	1906	69.5
1883	29.9	1891	46.8	1899	62.4	1907	73.5
1884	34.5	1892	47.8	1900	56.3	1908	70.8
1885	31.1	1893	50.2	1901	66.2	1909	74.1
1886	36.8	1894	60.1	1902	67.2	1910	76.1
1887	39.9	1895	55.4	1903	71.0		

TABLE XXI. (a)

PROPORTION OF TOTAL DEATHS OCCURRING IN HOSPITALS, NEW YORK CITY, 1898–1915.

Year.	Total Deaths.	Deaths in Hospitals.	Per Cent. Occurring in Hospitals.
1898	66,294	10,799	16.3
1899	65,343	11,749	18.0
1900	70,872	12,609	17.8
1901	70,720	13,343	18.9
1902	68,112	12,613	18.5
1903	67,864	14,695	21.6
1904	78,060	17,138	22.0
1905	73,714	18,103	24.6
1906	76,203	19,163	25.1
1907	79,205	21,444	27.1
1908	73,072	20,684	28.3
1909	74,105	21,451	28.9
1910	76,742	22,631	29.5
1911	75,423	23,466	31.1
1912	73,008	22,198	30.4
1913	73,902	22,788	30.8
1914	74,803	23,823	31.8
1915	76,193	25,095	32.9

(a) Figures kindly furnished the writer by the Bureau of Records, New York City. Hospitals, as here used, include only those institutions "which individuals enter direct from their homes," and not those "of which decedents were residents at the time of death."

Digitized by Google

TABLE XXII. (a)

PROPORTION OF DEATHS IN HOSPITALS TO TOTAL DEATHS, FORMER CITY OF NEW YORK, 1889–1897.

Year.	Total Deaths.	Deaths in Hospitals.	Per Cent. of Deaths Occurring in Hospitals.
1889	39,582	6,102	15.4
1890	40,103	6,310	15.7
1891	43,650	6,948	15.9
1892	44,329	7,681	17.3
1893	44,486	8,306	18.7
1894	41,175	7,742	18.8
1895	43,420	8,091	18.6
1896	41,622	8,530	20.4
1897	38,857	8,534	22.0

(a) Data derived from the successive *Annual Reports of the New York City Department of Health.*

TABLE XXIII. (a)

HOSPITAL BEDS TO 100,000 POPULATION, GERMAN EMPIRE, 1877–1910.

Year.	Hospital Beds per 100,000 Population.	Year.	Hospital Beds per 100,000 Population.
1877	166	1902	325
1879	175	1903	335
1882	182	1904	345
1885	201	1905	356
1888	224	1906	364
1891	246	1907	376
1894	266	1908	382
1897	273	1909	391
1900	295	1910	400
1901	310		

(a) Computed from *Stat. Jahrb. für d. D. Reich*, XXXVI (1915), population p. 2; hospital statistics p. 446.

TABLE XXIV.

PROPORTION OF DEATHS FROM CANCER AND FROM TUMOR TO DEATHS FROM BOTH CAUSES, UNITED STATES, 1880–1914.

Year.	Deaths from			Per Cent. Cancer.	Per Cent. Tumor.
	Cancer and Tumor.	Cancer.	Tumor.		
UNITED STATES (ENTIRE AREA).					
1880	14,849	13,068	1,781	88.0	12.0
1890	20,984	18,536	2,448	88.3	11.7
1900	32,902	29,475	3,427	89.6	10.4

Digitized by Google

UNITED STATES (REGISTRATION AREA). (a)

1900	19,962	19,381	581	97.1	2.9
1901	20,676	20,171	505	97.5	2.5
1902	21,306	20,847	459	97.8	2.2
1903	22,856	22,325	531	97.7	2.3
1904	23,839	23,395	444	98.1	1.9
1900–1904	108,639	106,119	2,520	97.7	2.3
1905	24,831	24,330	501	98.0	2.0
1906	29,490	29,020	470	98.4	1.6
1907	31,022	30,514	508	98.4	1.6
1908	33,923	33,465	458	98.6	1.4
1909	38,020	37,562	458	98.8	1.2
1905–1909	157,286	154,891	2,395	98.5	1.5
1910	41,592	41,039	553	98.7	1.3
1911	44,479	44,024	455	99.0	1.0
1912	46,908	46,531	377	99.2	0.8
1913	50,225	49,928	297	99.4	0.6
1914	52,756	52,420	336	99.4	0.6
1910–1914	235,960	233,942	2,018	99.1	0.9

(a) See comments, pp. 734–735.

TABLE XXV. (a)

DEATH RATES FROM CANCER AND FROM TUMOR IN THE REGISTRATION STATES OF 1900, FOR THE YEARS 1900-1915.

Year.	Population.	Deaths from			Death Rate per 100,000 Population.			Per Cent. of Deaths from Cancer.
		Cancer.	Tumor.	Both.	Cancer.	Tumor.	Both.	
1900	19,995,212	12,769	394	13,163	63.9	2.0	65.9	97.0
1901	20,408,866	13,438	332	13,770	65.8	1.6	67.4	97.6
1902	20,822,520	13,653	287	13,940	65.6	1.4	67.0	97.9
1903	21,236,174	14,650	355	15,005	69.0	1.7	70.7	97.6
1904	21,649,829	15,247	294	15,541	70.4	1.4	71.8	98.1
1905	22,063,484	15,983	318	16,301	72.4	1.4	73.8	98.1
1906	22,477,139	16,426	284	16,710	73.1	1.3	74.4	98.3
1907	22,890,794	17,324	278	17,602	75.7	1.2	76.9	98.5
1908	23,304,449	17,896	229	18,125	76.8	1.0	77.8	98.7
1909	23,718,104	18,830	227	19,057	79.4	1.0	80.4	98.8
1910	24,131,759	20,024	294	20,318	83.0	1.2	84.2	98.6
1911	24,545,413	20,596	203	20,799	83.9	0.8	84.7	99.1
1912	24,959,067	21,451	169	21,620	85.9	0.7	86.6	99.2
1913	25,372,721	22,497	119	22,616	88.7	0.5	89.2	99.4
1914	25,786,376	22,967	144	23,111	89.1	0.6	89.7	99.3
1915	26,200,032	24,061	138	24,199	91.8	0.5	92.3	99.4

(a) Population, deaths and death rates from cancer kindly furnished the writer in manuscript by the Bureau of the Census; figures for tumor compiled from successive volumes of *Mortality Statistics*.

Digitized by Google

TABLE XXVI. (a)

DEATH RATES FROM CANCER AND FROM TUMOR, MASSACHUSETTS, BY QUINQUENNIAL PERIODS, 1853–1912.

Period.	Mid-period Census Population.	Average Annual Number of Deaths from			Death Rates per 100,000 Population.			Per Cent. of Deaths from Cancer.
		Cancer.	Tumor.	Both.	Cancer.	Tumor.	Both.	
1853–1857	1,132,369	219	60	279	19.3	5.3	24.6	78.5
1858–1862	1,231,066	317	61	378	25.8	4.9	30.7	83.9
1863–1867	1,267,031	376	67	443	29.7	5.3	35.0	84.9
1868–1872	1,457,351	509	84	593	34.9	5.8	40.7	85.8
1873–1877	1,651,912	618	90	708	37.4	5.4	42.8	87.3
1878–1882	1,783,085	907	68	975	50.8	3.8	54.6	93.0
1883–1887	1,942,141	1,090	64	1,154	56.1	3.3	59.4	94.5
1888–1892	2,238,943	1,357	70	1,427	60.6	3.1	63.7	95.1
1893–1897	2,500,183	1,677	95	1,772	67.1	3.8	70.9	94.6
1898–1902	2,805,346	1,993	85	2,078	71.0	3.0	74.0	96.0
1903–1907	3,003,680	2,502	76	2,578	83.3	2.5	85.8	97.1
1908–1912	3,366,416	3,039	42	3,081	90.2	1.2	91.4	98.6

(a) Population from federal and state censuses; deaths from cancer and tumor from successive state Registration Reports.

TABLE XXVII. (a)

DEATH RATES FROM OLD AGE IN THE REGISTRATION STATES OF 1900, FOR THE YEARS 1900–1915.

Year.	Population.	Deaths from Old Age.	Death Rates from Old Age per 100,000 Population.
1900	19,995,212	10,015	50.1
1901	20,408,866	9,771	47.9
1902	20,822,520	9,309	44.7
1903	21,236,174	8,604	40.5
1904	21,649,829	8,702	40.2
1905	22,063,484	8,251	37.4
1906	22,477,139	8,432	37.5
1907	22,890,794	7,614	33.3
1908	23,304,449	6,943	29.8
1909	23,718,104	6,472	27.3
1910	24,131,759	6,321	26.2
1911	24,545,413	6,070	24.7
1912	24,959,067	6,426	25.8
1913	25,372,721	5,795	22.8
1914	25,786,376	4,901	19.0
1915	26,200,032	4,380	16.7

(a) Population figures from manuscript table kindly furnished by the Bureau of the Census; deaths from old age from the successive volumes of *Mortality Statistics*.

Digitized by Google

TABLE XXVIII. (a)

DEATH RATE FROM CANCER IN CITIES AND RURAL DISTRICTS OF THE REGISTRATION STATES OF 1900, FOR YEARS 1900–1915.

Year.	Population.		Deaths from Cancer.		Death Rate per 100,000 Population.	
	Cities.	Rural Districts.	Cities.	Rural Districts.	Cities.	Rural Districts.
1900	10,690,666	9,304,546	7,060	5,709	66.0	61.4
1901	11,105,137	9,303,729	7,717	5,721	69.5	61.5
1902	11,519,608	9,302,912	7,844	5,809	68.1	62.4
1903	11,934,080	9,302,094	8,498	6,152	71.2	66.1
1904	12,348,553	9,301,276	8,892	6,355	72.0	68.3
1905	12,763,026	9,300,458	9,323	6,660	73.0	71.6
1906	13,177,499	9,299,640	10,065	6,361	76.4	68.4
1907	13,591,972	9,298,822	10,544	6,780	77.6	72.9
1908	14,006,445	9,298,004	10,809	7,087	77.2	76.2
1909	14,420,918	9,297,186	11,530	7,300	80.0	78.5
1910	14,838,879	9,292,880	12,302	7,722	82.9	83.1
1911	15,202,813	9,342,600	12,760	7,836	83.9	83.9
1912	15,565,218	9,393,849	13,437	8,014	86.3	85.3
1913	15,927,982	9,444,739	14,042	8,455	88.2	89.5
1914	16,294,421	9,491,955	14,546	8,421	89.3	88.7
1915	16,662,676	9,537,356	15,208	8,853	91.3	92.8

(a) Data kindly furnished the writer in manuscript by the Bureau of the Census.

TABLE XXIX. (a)

DEATH RATES FROM CANCER IN CERTAIN EUROPEAN COUNTRIES CLASSIFIED BY CITY AND REST OF COUNTRY.

Year.	Population.		Deaths from Cancer.		Death Rates from Cancer per 100,000 Population.	
	Urban.	Rest of Country.	Urban.	Rest of Country.	Urban.	Rest of Country.
			ENGLAND AND WALES. (b)			
1881	5,058,823	20,987,319	2,898	10,644	57.3	50.7
1891	5,546,621	23,539,198	4,217	15,900	76.0	67.5
1901	6,141,048	26,471,086	5,553	21,934	90.4	82.9
1909	6,222,293	29,205,379	6,687	27,366	107.5	93.7
			SCOTLAND. (c)			
1881	740,890	3,001,674	424	1,490	57.2	49.6
1891	829,394	3,206,851	584	2,119	70.4	66.1
1901	1,092,515	3,386,550	864	2,798	79.1	82.6
1909	1,102,996	3,604,862	1,213	3,569	110.0	99.0

(a) Data from *Hoffman*, supplemented by material from *Falkenburg* and *March*.
(b) Cities are Birmingham, Bradford, Derby, Liverpool and London.
(c) Cities are Edinburg and Glasgow.

Digitized by Google

Norway. (d)						
1881	120,722	1,793,278	105	769	87.0	42.9
1891	153,887	1,842,113	147	1,077	95.5	58.5
1901	228,336	2,006,664	222	1,904	97.2	94.9
1909	239,704	2,098,296	239	1,987	99.7	94.7
Netherlands. (e)						
1881	683,889	3,403,445	475	1,878	69.5	55.2
1891	890,775	3,702,379	743	2,905	83.4	78.5
1901	1,193,689	4,023,554	1,104	3,790	92.5	94.2
1909	1,366,862	4,453,613	1,385	4,627	101.3	103.9
Austria. (f)						
1881	976,535	21,232,465	1,162	8,202	119.0	38.6
1891	1,683,096	22,312,904	2,018	11,009	119.9	49.3
1901	2,065,344	24,113,412	2,462	16,692	119.2	69.2
1908	2,416,104	25,427,421	3,207	18,973	132.7	74.6

(d) City is Christian.
(e) Cities are Amsterdam, the Hague, Rotterdam and Utrecht.
(f) Cities are Prague and Vienna.

TABLE XXX. (a)

DEATH RATES FROM CANCER, CLASSIFIED BY COLOR OF DECEDENT, IN THE REGISTRATION STATES OF 1910, FOR THE YEARS 1910–1915.

Year.	Population.	Deaths from Cancer.	Death Rate from Cancer per 100,000 Population.
Total.			
1910	47,807,766	36,364	76.1
1911	48,756,771	37,269	76.4
1912	49,595,220	39,279	79.2
1913	50,493,332	41,430	82.1
1914	51,455,677	42,706	83.0
1915	52,357,716	44,440	84.9
White.			
1910	46,425,660	35,638	76.8
1911	47,344,587	36,432	77.0
1912	48,153,883	38,424	79.8
1913	49,025,379	40,490	82.6
1914	49,941,006	41,716	83.5
1915	50,845,959	43,400	85.4
Colored.			
1910	1,382,106	726	52.6
1911	1,412,184	837	59.3
1912	1,441,337	855	59.3
1913	1,467,953	940	64.0
1914	1,514,671	990	65.4
1915	1,511,757	1,040	68 8

(a) Figures for 1910–1913 from a manuscript table kindly furnished the writer by the Bureau of the Census.

Digitized by Google

TABLE XXXI. (a)

DEATH RATES FROM CANCER CLASSIFIED BY SEX AND AGE, AND RATE OF DECENNIAL INCREASE, REGISTRATION STATES OF 1900, FOR 1900 AND 1910.

Age.	Males.			Females.		
	Death Rate from Cancer per 100,000 Population.		Rate of Increase 1900–1910.	Death Rate from Cancer per 100,000 Population.		Rate of Increase 1900–1910.
	1900.	1910.		1900.	1910.	
25–34	9.1	9.5	4	18.9	21.9	16
35–44	28.1	33.0	17	78.8	88.9	13
45–54	83.5	106.7	28	196.6	230.7	17
55–64	189.3	272.0	44	330.5	411.3	24
65–74	377.4	493.6	31	461.6	616.2	33
75+	516.9	693.7	34	609.1	867.8	42

(a) From a manuscript table kindly furnished the writer by the Bureau of the Census.

TABLE XXXII. (a)

DEATH RATES FROM CANCER PER 100,000 POPULATION CLASSIFIED BY SEX AND AGE, MASSACHUSETTS, 1860–1910.

Age Period.	1860.	1870.	1880.	1890.	1900.	1910.
			Males.			
30–39	6.6	6.9	11.9	7.1	16.9	17.9
40–49	20.4	27.5	39.0	29.3	62.5	61.4
50–59	63.4	83.4	99.7	118.2	145.6	161.6
60–69	105.7	154.2	168.7	267.1	277.6	415.9
70–79	166.6	267.2	408.3	332.8	497.1	658.4
80+	349.7	354.8	491.1	331.7	655.4	955.1
			Females.			
30–39	17.1	23.0	46.3	45.5	38.1	48.5
40–49	79.5	84.8	116.0	138.2	149.2	160.5
50–59	145.2	139.7	219.4	261.8	284.5	317.9
60–69	162.0	216.9	282.7	387.6	451.4	558.5
70–79	202.5	254.8	371.9	527.3	555.2	817.7
80+	341.4	379.4	320.2	438.2	680.1	861.4

Digitized by Google

PER CENT. OF DECENNIAL INCREASE IN DEATH RATE.

Age Period.	1860–1870.	1870–1880.	1880–1890.	1890–1900.	1900–1910.
			MALES.		
30–39	4.5	72.5	−40.3 (b)	128.0	5.9
40–49	4.2	41.8	−24.9 (b)	113.3	−1.8 (b)
50–59	31.5	19.5	18.6	23.2	11.0
60–69	45.9	9.4	58.3	3.9	49.8
70–79	60.4	52.8	−18.5 (b)	49.4	32.5
80+	1.5	38.4	−32.5 (b)	97.6	45.7
			FEMALES.		
30–39	34.5	101.3	−1.7 (b)	−16.3 (b)	27.3
40–49	6.7	36.8	19.1	8.0	7.6
50–59	−3.8 (b)	57.1	19.3	8.7	11.7
60–69	33.9	30.3	37.1	16.5	23.7
70–79	25.8	46.0	41.8	5.3	47.3
80+	11.1	−15.6 (b)	36.8	55.2	26.7

(a) Population figures used in computing these rates are those of the Federal censuses; deaths, those in the successive *Massachusetts Registration Reports*.
(b) Decrease.

TABLE XXXIII. (a)

DEATH RATE FROM APPENDICITIS IN THE REGISTRATION STATES OF 1900, FOR THE YEARS 1900–1915.

Year.	Population.	Deaths from Appendicitis.	Death Rate per 100,000 Population.
1900	19,995,212	1,763	8.8
1901	20,408,866	1,794	8.8
1902	20,822,520	1,825	8.8
1903	21,236,174	2,031	9.6
1904	21,649,829	2,224	10.3
1905	22,063,484	2,256	10.2
1906	22,477,139	2,318	10.3
1907	22,890,794	2,326	10.2
1908	23,304,449	2,396	10.3
1909	23,718,104	2,550	10.8
1910	24,131,759	2,685	11.1
1911	24,545,413	2,783	11.3
1912	24,959,067	2,743	11.0
1913	25,372,721	2,940	11.6
1914	25,786,376	3,149	12.2
1915	26,200,032	3,221	12.3

(a) Data kindly furnished the writer in manuscript by the Bureau of the Census.

Digitized by Google

SPECIAL MEETING OF THE AMERICAN STATISTICAL ASSOCIATION.

About 50 members and guests attended a special meeting and dinner of the Association held at the Aldine Club, New York City, Friday evening, April 20, 1917. Prof. Allyn A. Young, President of the Association, presided. The subject of the meeting was the value, especially to business interests, of federal census statistics, and possible improvements for the next decennial census in 1920. Prof. Robert E. Chaddock, Assistant Secretary of the Association, arranged the very interesting program.

The discussion was opened by W. A. Hathaway* of the American Telephone and Telegraph Company, who first briefly sketched the growth of the census from 1790 (the first census) to the present time, and then suggested several improvements in both scope and methods. In speaking of the next census, he discussed the desirability of: (1) keeping politics out of the census work, (2) certain changes in the schedules and classifications for manufactures, (3) a greatly simplified schedule for agriculture, (4) earlier publication of results, especially "Preliminary" totals, and (5) certain *annual* statistics of agriculture and manufactures, particularly the latter.

Various aspects of the census schedules and publications were then briefly discussed by the following speakers:

Colonel George Pope (National Association of Manufacturers)—emphasized the great value of census statistics to manufacturers, and pointed out the possible advantages of greater detail in industry classifications, and the desirability of collecting statistics at more frequent intervals.

O. P. Austin (National City Bank)—quoted census figures to show the need of annual censuses of agriculture and manufactures.†

* Mr. Hathaway was Chief of the Division for Agriculture and Chief Clerk of the Bureau during the 13th U. S. Census, 1910. (Editor's note.)

† His paper is printed in full elsewhere in this issue.

Digitized by Google

Osmund Phillips (Times Analist)—suggested that: (1) the effect of higher prices on the increase in value of products of manufactures be stated, (2) the hours of labor required to produce the products reported for each industry be shown, (3) a classification of the size (value of products) of establishments by ownership be added, (4) the per capita consumption be computed by industries, and (5) the number of persons employed and value of products be secured each month or quarter from 1,000, or so, representative manufacturers, and these totals be used as a current index of industrial conditions.

E. A. Goldenweiser (Office of Farm Management of United States Department of Agriculture)—stated that the Department of Agriculture was coöperating with the census in drafting the agricultural schedule for 1920. He distributed copies of a tentative schedule for criticism.

M. C. Rorty (American Telephone and Telegraph Company)—made some very interesting observations on statistics of incomes, with especial reference to the possibility of securing trustworthy data without a complete enumeration.*

Frank Green (Bradstreets Commercial Agency)—called attention to the fact that distribution (wholesale and retail trade) was about the only line of business and industry not now covered by a federal census, and suggested that the government might well consider extending its statistical activity to cover this field. He pointed out some of the many ways in which such data could be used to supply the missing link in the statistical chain, and make it possible to correlate much unrelated data.

F. H. Kenny (World Almanac)—expressed great appreciation of the service that census statistics were rendering to all classes of business interests. He strengthened the arguments in favor of annual censuses by stating that the majority of requests which he received were for *later* figures.

G. H. Van Buren (Metropolitan Life Ins. Co.)—reported that the American Public Health Association was now studying the population schedules with a view to making suggestions for the next census.

* His paper is printed in full elsewhere in this issue.

Digitized by Google

The President called upon Dr. R. P. Falkner* (Alexander Hamilton Institute) and Dr. J. A. Hill (Bureau of the Census) to close the discussion. Dr. Falkner elaborated on some of Mr. Rorty's observations on income statistics, and commented on several aspects of census work. Dr. Hill expressed appreciation, on behalf of the census, of the friendly suggestions and criticisms put forth by the several speakers. He said the census officials welcomed this sort of help and coöperation, and expressed the belief that much good would result from the discussion of the evening.

* Former Assistant Director of the Census.

Digitized by Google

SHOULD WE HAVE AN ANNUAL CENSUS OF AGRICULTURE AND MANUFACTURES?*

By O. P. Austin,
Statistician, The National City Bank of New York.

The question should we have a more frequent census of agriculture and agricultural products should unquestionably, in my opinion, be answered in the affirmative. Whether the more explicit question, should we have an *annual* census of agriculture should be answered in the affirmative there might be a difference of opinion. At present we have a decennial census of population, mining, and forestry, and a quinquennial census of manufactures, and it seems to me that nobody can doubt the advisability of having a census of agriculture even more frequently than that of manufactures, while the ideal plan would of course be an annual census for agriculture since a large proportion of the agricultural productions to be enumerated are annual in their recurrence.

And there are special reasons just at this time for feeling that agriculture is entitled to very close attention statistically as well as otherwise. We have reached a stage in our agricultural production in which scientific farming is of the utmost importance. We who, until within a comparatively recent time, were looked upon as the world's great purveyor of foodstuffs and manufacturing material have suddenly found that we have little of foodstuffs to spare for the outside world, and that we are ourselves consuming an increasing percentage of our manufacturing material. In the decade ending with 1900 we exported 36 per cent. of our wheat, but from that time exports declined, and in the 13 years from 1900 to the beginning of the war, we exported but 19 per cent. of our product, and in several years but 12 per cent. With the pressing demands of the war, however, the share of our products exported immediately increased and for the entire war period is about 29 per cent. of the total product. This increase in the percentage of our crop exported has, however, been at the

*Address delivered before the American Statistical Association at the Aldine Club, New York, April 20, 1917.

Digitized by Google

expense of our consumers, as is seen in the unprecedentedly high prices which now prevail, and the public anxiety as to sufficiency of supply pending arrival of the new crop. It is thus especially important, now that our population has reached a point at which it consumes most of our agricultural supplies, to intensify the product wherever possible, and to require our farms to supply as large a production per acre as do the much older lands of Europe where the production per acre averages about double that with us. To do this we should have all the information possible thereon, and this should include much more frequent official records of production than we have at the present time.

I use this term "records of production" because I want to distinguish between our present *decennial record* of production and the *annual estimate* of production supplied by the Department of Agriculture. In doing so I desire to make it perfectly clear that my suggestions are not intended as in any sense a criticism of or reflection upon the methods of the Department of Agriculture. That department, in fact, is performing a valuable service in its estimates and statistics of production, which it issues annually in certain lines, and from month to month during the crop year. But the facilities with which it operates and is compelled under present conditions to operate, and the limited appropriations which it has for performing that service cannot compare with those of a census, nor can anybody expect that the results will equal those of a census with a large corps of salaried enumerators covering the entire country in much greater detail than is possible by the Department of Agriculture under its present system of estimates by voluntary reports. In proof of this assertion it is only necessary to compare the methods of the two organizations of the Department of Agriculture, and the Census Bureau as stated by those respective organizations.

In the latest issue of the Monthly Crop Report, published by the Department of Agriculture, I find a detailed statement of the method by which the department collects the material upon which it bases its crop reports and estimates. Whether or not this statement was published at this particular moment with the purpose of supplying definite information for this

Digitized by Google

particular occasion, it at least does come to us at an opportune moment. It says:

The crop reports of the Department of Agriculture are not based upon actual enumeration of the total acreage planted of the different crops throughout the country, but are estimates, the accuracy of which rests upon the assumption that the average crop conditions existing at 50,000 separate points including every agricultural county of the United States, will be substantially the same as the conditions for all of the land under cultivation in this country. In other words, the government reports rest upon the scientific basis of the law of average. Therefore it is not necessary that voluntary crop reporters should report for an absolute area, rather it is important that they should give figures for that section only with which they are personally familiar by their personal observations and conversations with dependable neighbors. If each reporter will report accurately for his own neighborhood, the average for the state cannot fail to be almost absolutely correct. While the township is the ideal unit for reporting, these township reporters should not feel bound to report for the entire area embraced in the township, because experience shows that on the average the conditions existing in that part of the township under his immediate observation will faithfully reflect the condition over the entire township. County reporters are expected to report for the county as a whole upon practically the same principle, their reports being based upon their personal observations, reinforced by reports to them from different portions of the county. The average obtained from the reports from different sources, township reporters, county reporters, state field agents, and special lists of reports, are usually in exceedingly close agreement, and for important crops rarely varying more than 5 between the highest and the lowest figure.

Here we have a definite statement from the Department of Agriculture of the method by which it prepares the estimates of crop and other conditions, and it is only fair in considering this question of whether we should have more frequent census reports of agriculture to contrast the system of the Department of Agriculture with that of the Census Bureau. To do this I quote to you from the 1910 Report of the Census of Agriculture, the method by which its statistics were compiled. First, the law providing for the census of agriculture required that schedules relating to agriculture should include the name of the occupant of each farm, and among other requirements,

Digitized by Google

the value of live stock, the acreage of crops planted during the year of the enumeration, and the quantity and value of crops and other farm products in the year next preceding the enumeration. Each census supervisor was directed to employ suitable persons as enumerators who were respectively required to "visit personally each dwelling house in his division and each family therein, and by inquiry made of the head of each family, or of the member thereof deemed most competent and most trustworthy, obtain each and every item of information required by this act, and in case no person shall be found at the usual place of abode of such family competent to answer the inquiries, then the enumerator shall obtain the required information as nearly as may be practicable from families or persons living in the neighborhood of such place of abode."

It will be seen from this extract from the law that enumerators carry a uniform schedule to *every farm in the United States*, and obtain from the occupant of the farm, or if this is not possible, from others best-equipped to supply the information, the data upon which the census report of production was made. The Department of Agriculture statement just read to you states that its estimates are based upon the opinion of 50,000 voluntary reporters. The census states that 5,795,000 farmers reported upon the schedules placed before them by the census enumerators, and while it seems reasonable to assume that the judgment of 50,000 Department of Agriculture reporters was good, it would appear that the judgment of the 5,795,000 farmers who actually produced the grain and live stock reported upon would be materially better in shaping up measurements or estimates of production.

That there is a marked discrepancy between the Department of Agriculture estimates, and the census count, will be evident from a comparison of a few of the statements of these two organizations relating to production in census years. In 1879 the Department of Agriculture figures as published by the Statistical Abstract of the United States, estimated the corn crop of the country 181,000,000 bushels below the census count for that year. In 1899 its wheat figures were 112,000,000 bushels below those of the census, and those of corn were 588,000,000 lower than those of the census. In 1909 its corn

Digitized by Google

figures were 220,000,000 bushels higher than those of the census, and its wheat figures, 54,000,000 above the census record. In the estimates of live stock there were greater discrepancies, supposed to be due in some degree to the fact that the census figures related to a different period of the year than those of the department, and included many young animals not in existence at the date of the department estimate as of January 1. The Agricultural Department, however, seems, after years of wide differences in the two reports to have concluded to give closer recognition to the census figures for in its published estimates of farm animals the figures of 1901, the year following the census of 1900 show a sudden and enormous increase, and are accompanied by a statement that "in estimates of 1901 and subsequent years due weight has been given to the figures of the census." The "weight" given to the census figures must have been very appreciable, since its estimates of cattle in 1901 are 42 per cent. greater than those of 1900, sheep 44 increase, swine 53, mules 37, and horses 22 per cent. greater in the 1901 figures than those of 1900, while the estimated value of all farm animals in 1901 was 36 per cent. greater than for 1900. By 1910, however, the department estimates seemed to have again required readjustment, for its 1911 figures showed a sudden increase of 39 per cent. in the estimated number of swine as compared with the year immediately preceding, while the group "oxen and other cattle" showed a reduction of 16 per cent. And all of these great changes it must be remembered occurred in the year immediately following that in which the census figures became available.

I am not making these comparisons with the purpose of criticising the figures of the Department of Agriculture, to which we are all under obligations for the results of the system by which it obtains with much greater frequency than the census the individual opinions of such a large force of observers in every part of the country, a service in which it produces prompt and valuable results, but I am only attempting to emphasize the fact that Congress is neglecting a duty to the great agricultural as well as economic interests of the country in omitting to require its great counting machine,

Digitized by Google

the Census Bureau, to apply from year to year to this most important of our industries the systematic and detailed study which it now applies only once in ten years.

Our manufacturing industry is at this time also very much in need of an additional service from the Census Bureau. For many years the decennial census of manufactures served a fairly satisfactory purpose, but with the rapid increase in manufacturing in the United States, and the equally rapid growth in the exportation of manufactures, there came a time when a more frequent census of manufactures was demanded by those engaged in that industry, whether as relates to capital, management, or employment. As a consequence, we have had since 1900 quinquennial censuses of manufactures, whereas prior to 1900 they were but decennial. And the time has now come, in my opinion, when the censuses of manufactures should be even more frequent. Down to 1900 the natural products, foodstuffs, and manufacturing material formed about two thirds of our exports, which had only just crossed the $1,000,000,000 line. As we entered the twentieth century there came a sharp change in the character of our exports; foodstuffs fell off amazingly, so that they formed in the year before the war only 18 per cent. of the total against 40 per cent. in 1900 and 56 per cent. in 1880. Conversely, manufactures, which formed in 1880 but 15 per cent. of the total exports were in 1900 35 per cent. and in 1913 practically 50 per cent. while since the beginning of the war, they have formed a much larger proportion, being 66 per cent. in 1916. It is perfectly apparent that we shall have in future, when conditions after the war return to normal, very little of foodstuffs to spare to the outside world, and that we are ourselves consuming an increasing proportion of the one class of manufacturing material which we have to sell, namely, cotton. All of this means that our future reliance for the growth of our exports must be manufactures. They have grown in the export trade, including in this figure finished manufactures and those for further use in manufacturing, from $180,000,000 in 1890 to $485,000,000 in 1900, $767,000,000 in 1910, and $1,185,000,000 in 1913, while with the abnormal demands of the world upon us, they grew to over $3,500,000,000 in the calendar year 1916, having

Digitized by Google

thus been in the calendar year 1916 more than three times as much as in the fiscal year 1913.

This growth in exportation of manufactures from $180,000,000 in 1890 to $1,185,000,000 in 1913 was normal, and while the enormous growth from 1913 to the end of the calendar year 1916 was of course abnormal, it does illustrate the possibilities of our manufacturers in producing for foreign markets, and suggests that they may be able in the future to assume the responsibility of maintaining the rate of growth in our total exports which occurred chiefly in foodstuffs during the earlier period in our history. And have we not good reason to believe from the lessons of the last three years that our manufactures are capable of assuming this new responsibility of increasing their share in the growth of our export trade? Without a very rapid growth in exports of manufactures our rank as an exporting nation must decline. And it seems to me that we may properly assume that a manufacturing industry which could under normal conditions increase its exports from $180,000,000 in 1890 to $1,185,000,000 in 1913, and follow this with the enormous increase which we have witnessed during the war period, may be relied upon to assume the responsibility for a continuation of the growth of our entire exporting trade despite the fact that exports of foodstuffs must decline in the immediate future.

To enable the manufacturing industries to assume this new duty, we owe it to them, in my opinion, to give them much more frequent official information regarding the manufacturing activities of the entire country. Each recurring census shows to each manufacturer of the entire country the amount of capital invested, the number of workmen employed, the amount of material purchased, and the value of output in the line or lines of industry in which he is individually engaged. That this information must be of great value to every manufacturer in broad studies of the development of the manufacturing industry, and especially his own particular line, goes without saying, and if such information is valuable to him when it comes on decennial or quinquennial occasions, how much more valuable would it be if he could have it annually?

I do not mean by this that we should take an annual census

Digitized by Google

of manufactures in the very great detail in which the quinquennial census is now made, but for a limited number of lines, those which form the large proportion of the output of our great manufacturing industry, an annual census could be taken at a comparatively small cost, and if in selecting these industries special attention were given to those in which we desire to increase our sales abroad it would prove of very great service to our export trade and to our manufacturing industry as a whole.

Digitized by Google

INCOME STATISTICS.*

By M. C. Rorty.

In these troubled days, when almost every element of governmental policy, and even the fate of whole peoples, hinges upon the completeness and reliability of statistics, it is not necessary to argue the practical value that may lie in knowing the basic facts in regard to our national economy.

The question is not whether we shall have statistics, but how the statistics that the inexorable logic of events has forced, and will force, us to set up may be made most accurate and most pertinent to our national purposes.

Income statistics have their immediate use in determining and controlling state and federal policies in respect to income taxes. Such taxation, in one form or another, is apparently with us to stay, and the needs of the taxing authorities are, in themselves, sufficient justification for a thorough analysis of the facts.

This need, while the most obvious, is not, however, the most important. Opinions may differ widely as to the ultimate outcome of the present world upheaval, but skilled observers are almost unanimous in predicting radical internal changes of a semi-socialistic type in the nations most directly involved. Such changes can hardly fail to be reflected by a period of political unrest and political experimenting in our own country, and this experimenting, directly or indirectly, must center, in the end, around that ever present and greatest of problems —the family income.

It is idle to ask whether the average voter may be trusted to give sane answers to questions of government ownership, of land nationalization, of labor legislation, of social insurance, and many others that will arise. The sufficient answer is that he must be trusted, since he has the power and the will to decide.

The second and greatest purpose of income statistics is, therefore, to furnish a sound basis for popular consideration of social and political problems.

* Paper read before the American Statistical Association at New York, April 20, 1917.

Digitized by Google

From the standpoint of taxation, the special requirement is for knowledge of the upper range of incomes, including those subject to proposed as well as present taxes. Within this range, which may be assumed to extend from $1,000 or $1,500 per annum upward, it is rather obvious that no census return, for any given individual, is apt to be more accurate than a sworn statement filed with and checked by the Collector of Internal Revenue. This being the case, the next problem is to determine how far below the range of actual taxation it may be desirable to extend the requirement for sworn returns, as a check against omissions and to provide advance indications as to the results that might be obtained from any proposed lowering of the limits of exemption. The answer will depend, to a considerable extent, on the general information that is available as to the lower range of incomes and the difficulties that are experienced in reconciling the income tax returns with such general data in the range of overlap.

A practical solution at this time might be to require returns from all persons having incomes within, say, $500 of the lower limit of exemption, and to require returns to be filed, in addition, regardless of income, by all heads of families and others occupying living quarters renting for, or having a rental value of, more than a specified amount. The latter provision would very considerably simplify the checking of the total number of returns filed, and would be in line with common knowledge that 10 per cent. or more of the average family income is paid in rent or its equivalent.

With these or other similar modifications of the existing laws in effect, the only remaining problem, from a statistical standpoint, would be that of determining the analysis that should be given to the returns as received. This matter is now being studied by a joint committee of this Association and the American Economic Association, in coöperation with the Commissioner of Internal Revenue, and there is every reason to believe that, within the limits fixed by permissible expense for tabulation, a very satisfactory analysis of income tax returns will soon be available.

Considering next the use of income statistics as the basis for popular and legislative consideration of social and political

Digitized by Google

questions, an entirely different statistical viewpoint is required. The problem in this respect is almost as largely one of publicity and education as it is of economic data.

The average worker today, as the most superficial inquiry will show, has a conviction, deeper and more firmly seated than his theology, that, if justice were done, he would receive at least two or three times his present wage. This conviction is also largely, if not so intensely, shared by educated and professional men of all classes; and the most curious feature of the whole situation is that, while each employer can prove that he, personally, is not a party to such injustice, a large part of employers themselves firmly believe that "big business," or the "money trust," or the "system" is at fault, and that they, as well as their employees, are being systematically robbed through some hidden kink in the general scheme of things.

It is this prevalent sense of a great social injustice that today offers the most fertile field for the efforts of the extremist and the I. W. W. agitator, and that a true and widespread knowledge of income distribution will have its greatest value in tempering to a point where sounder and more constructive counsels may be heard.

The differences between popular impressions and the real facts are, however, so great that the immediate need is not for an extreme elaboration of census procedure or for extreme exactitude of statistical results, but rather for breadth, impartiality, and comprehensiveness of viewpoint, and flexibility of approach. It is much more important, from the standpoint of the major uses of national statistics, that they shall tell a complete, coherent, and essentially truthful story of our national life, than that they shall be 100 per cent. exact in all details.

It is this point of view, among others, that led to the adoption, by the Chamber of Commerce of the United States, at its recent annual meeting, of resolutions in favor of the appointment of a Federal Statistical Commission to correlate and weld into a coherent whole the unrelated, although otherwise excellent, statistical work of the several federal departments, bureaus, and commissions.

Digitized by Google

Should such a commission be appointed, I have no doubt that it would assume, as a part of its task, the setting up, through the proper federal departments, of statistics covering the whole range of individual and family incomes, rather than the taxable range alone; and these statistics might properly be supplemented by a distribution of the total of such incomes according to basic sources,—wages and other returns for personal service, land rents, interest, and profits* in excess of a normal interest rate.

Such an analysis would have a practical value that will be appreciated at once by the members of this Association, and, in time, should be quite as fully appreciated by all those less technically trained who are interested in an orderly national evolution.

In approaching this analysis, I must differ, however, with Dr. King and others who would apply strict census methods throughout. My impression is that the final accuracy of results will be greater, and the expense much less, if census methods, as to certain basic elements in the determinations, are supplemented by skilled sampling processes for the establishment of related figures. For example, as to the element of land rents, the use of existing figures as to assessed valuations, in combination with skilled local estimates of the relation of site rents to such valuations, should give much more accurate results than could be obtained by any direct process.

Similarly, as to profits and interest charges in trade and industry, direct census methods might give much less exact results, even as to actual investment, than would be obtained by applying properly determined constants to basic figures as to average annual sales or output, number of employees, etc., for each class of establishments. Profits, in particular, vary so widely from year to year that only studies covering long periods, such as those recently made public by J. E. Sterrett, may safely be used, and the application of such long term

* In making this distribution, care should be taken to avoid combining "profits" and the "entrepreneur's wage." Profits should be defined as the excess of capital return over a normal interest rate, or, perhaps, over the average capital return in the marginal, or price fixing, group of establishments. The entrepreneur's wage should be treated simply as one class of returns for personal service. This distinction is particularly necessary in order to avoid giving false impressions as to social possibilities through profit sharing with employees, and might be carried a step further by making a distinction between profits of corporations and profits of individual entrepreneurs.

Digitized by Google

percentages to properly classified investment values might give more accurate general results than could be obtained by summing a great mass of individual returns in which the errors might readily be cumulative rather than compensatory. In this connection, it is especially important to note the vital distinction between interest and profits distributed in cash to individuals, and profits accumulated in a growing business or paid to other corporations. It is only interest and profits distributed in cash to individuals that become income of individuals. Undistributed profits may ultimately increase the volume of distributed profits and the total of individual incomes, but very frequently they become lost through the accidents of business before distribution is possible.

Considerations similar to the preceding hold true of the lower range of family incomes. Streightoff has suggested that the closest approximation as to city dwellers might be obtained through a census of actual and equivalent family rents. Such studies are privately made on an extensive scale as the basis for certain classes of engineering work and their practicability has been fully demonstrated. The range and distribution of family incomes associated with any given house rent undoubtedly varies in different communities, but the combined expense of a house rent census and the required special determinations should be very much less than that involved in any direct investigation; and the final results should be much less subject to systematic and cumulative errors. Figures as to house rents already at hand indicate that, within certain easily established classes of cities, there is so great a uniformity of distribution of families, in relation to either the median or the average rent, that it approaches an absolute law.

A further source of information as to individual incomes lies in the payrolls of the larger employers. These indicate, of course, earned incomes only, and cover only a fraction of the population, but I have reason to suspect that, under proper statistical analysis, they will reveal a law of distribution of such incomes, in relation to the average or median, that may be of great service in checking other figures.

All of the preceding are, of course, nothing but rough suggestions as to sources and methods. They may serve, how-

Digitized by Google

ever, to emphasize the fact that, for the broader uses of national income statistics, the first and most essential thing to have is what an eminent British statistician has aptly described as a "Central Statistical Thinking Office," and not a mere expansion and elaboration of uncorrelated statistical effort.

With this intelligent planning of our national statistical program we will need, also, an intelligent publicity, to which your own organization may contribute not a little.

A speaker at one of your dinners said, not long ago, that he had no interest at all in statistics, but was forever hunting for *the* statistic that had the punch and would carry conviction.

Let me conclude by giving rough illustrations of this point out of my own very sketchy study of income statistics.

First, as to earned incomes, large payrolls show that, if the excess of all salaries above $5,000 per annum were distributed among those getting less than $5,000 per annum, the average increase in pay in the lower range would be only 3 cents per day, and this increase would be only 11 cents per day if the limit were fixed at $1,500 per annum. War conditions may have shifted these figures somewhat, but cannot have changed the essential fact that no appreciable gain to the average worker can result from any conceivable evening up of earned incomes. An interesting side light on this point is that the average worker may doubt the propriety of some of the higher earned incomes, but, if questioned as to the relative industrial values of different classes of workers in the lower range, will very commonly indicate a greater spread in such values than is reflected in the average payroll.

Second, as to urban house rents, the average rent in typical large cities is only 20 per cent. or, say, $3.50 per month, higher than the median rent, and, if all sums now paid in excess of 150 per cent. of the average rent were distributed among families in the lower rental range, they could increase their present payments by less than 20 per cent.

Third, as to profits in trade and industry, the average real investment per employee is probably under rather than over $3,000, with book profits and interest charges of perhaps 14 per cent., and an actual cash distribution of perhaps 10 per cent. The excess of 4 per cent. above a normal 6 per cent.

Digitized by Google

interest rate corresponds to not more than $120 per annum, or somewhat less than 40 cents per day per employee.

These may sound like "stand pat" arguments, but they are indicative of facts which conservatives and radicals alike must face. It is reasonably clear, in any case, that the problem of a more equitable distribution of individual incomes, whatever that problem may be, lies in the unearned rather than the earned element, and is closely associated with the concentration of income producing wealth.

Digitized by Google

REVIEWS AND NOTES.

Social Diagnosis. By Mary E. Richmond, Charity Organization Department, Russell Sage Foundation, New York, 1917, pp. 511.

Statistics of social difficulties and of social needs are conditioned by the technique of social diagnosis. Social statistics are dependent upon social diagnosis in much the same way that public health statistics are primarily affected by the methods and terminology of physical and other forms of medical diagnosis.

Miss Richmond in her book outlines the processes which led to social diagnosis. These are, first, the collection and careful scrutiny of evidence rising out of the social worker's relations with his client, with the client's family, and with sources of information outside the family group; and second, the reflective comparison and interpretation of the evidence from these sources. The older and narrower process of "investigation" in social case work, with its loose technique and its besetting sin of almost exclusive emphasis upon the *gathering* of evidence to the detriment of essential matters of comparison and interpretation, is properly subordinated in this book.

The author considers first the nature and uses of social evidence as a basis for diagnosis. She outlines the logical tests for meeting in a practical manner the "enormous difficulty of making sure that the facts are as stated to us by others . . . and of avoiding that 'instinctive theorizing' whence the fact looks to the eye as the eye likes the look." The practicing social statistician will appreciate to the full the urgency of recording as exact a definition as possible of a given social situation and of its proximate and remote relations and causes. There is a commendably thorough discussion of the definitions bearing upon social evidence, of the terms most frequently used, and of the types of evidence, *i. e.*, whether real evidence, testimonial evidence, or circumstantial social evidence. Each of these types of evidence is examined in close detail. Testimonial evidence is especially well developed. A concluding chapter discusses inferences from the data collected, the risks involved in drawing conclusions from the facts at hand and from the social worker's particular state of mind.

The second section of the book discusses the detailed processes leading to diagnosis, and includes very extended technical comment on the scope and method of the interview with the client, with members of the family group, consultation of outside sources of information, the sources and validity of medical, legal and other records. A concluding chapter of the second part of the work defines the principles of criticism and comparison of the collected material, and the statement of a diagnosis.

Statisticians reading Miss Richmond's book, will recall the difficulties and limitations encountered by such pioneers as Amos Griswold Warner, Frederick H. Wines, and John Koren in the collection of social statistics in America. So many of these difficulties arose from the lack of any uniformity of methods of collecting and recording the basic data. One of the

Digitized by Google

effects upon American social statistics which should follow the introduction of works such as Miss Richmond has made available for social workers, should be the marked improvement of charity organization and other philanthropic agency records. Improvement of the statement of social difficulties upon the case records should result in better composite pictures of the social situation through the agency of the statistical method.

There is already a noteworthy precedent in the statistical study of social case work and of its results in the recently published report of the municipal court of Philadelphia for 1915. The methods employed by Dr. Louise Stevens Bryant in the compilation of these records for Philadelphia, if employed in studying the data which will undoubtedly result from employing Miss Richmond's criteria of social diagnosis, should provide for the first time in America the kind of social statistics which Professor Warner and the social workers and statisticians who followed him have defined as necessary for a critical analysis and an impartial survey of practical philanthropy.

New York City. E. W. Kopf.

Appendix A. Vol. I. Census of the Commonwealth of Australia. The mathematical theory of population, of its character and fluctuations, and of the factors which influence them, being an examination of the general scheme of statistical representation, with deductions of necessary formulae; the whole being applied to the data of the Australian census of 1911, and to the elucidation of Australian population statistics generally. By G. H. Knibbs, C.M.G., F.S.S., F.R.A.S., etc., commonwealth statistician. Published under instructions from the Minister of State for Home and Territories, Melbourne.

It is rare that a government census publication takes the form of this latest volume by Mr. Knibbs. As a rule census offices are so busy, collecting material, tabulating it, and getting out a brief analysis accompanied by a few charts, that they have no time to enter upon a scientific discussion of the uses which may be made of the volume. This is really a treatise on population statistics making use of the Australian figures primarily. In many cases the statistics of other countries are introduced for purposes of comparison. The first six chapters, covering about 85 pages, are devoted to a discussion of the formulas which may be used in determining the fluctuations and probable future course of populations. Then follows an interesting chapter upon graphics and the methods employed to smooth curves. After this, a discussion of population by sex and age, of natality, marriage rate, fecundity, mortality, and migration follows. An immense amount of work has gone into the preparation of this volume and the chart work is extremely clever and interesting. One would search long to find a better piece of analytical work than this. The census office of Australia is to be congratulated upon publishing such a volume and Mr. Knibbs deserves our thanks in showing the quality of work which can be done by its census office in addition to the details of collecting and compiling the basic figures.

Yale University. Wm. B. Bailey.

Digitized by Google

REED ON THE COEFFICIENT OF CORRELATION.

Since the development of the modern higher statistical methods has been largely in the service of biology, it is only natural that those who adopt the biometrician's formulae without a first hand acquaintance with the extensive biological literature should often labor under the disadvantage of using methods of calculation which are not the best suited to the practical needs of their work. If the methods adopted really give accurate constants, the use of more laborious routine processes is merely the misfortune of the individual worker. When, however, the beginner attempts, as is surprisingly often the case, to elucidate the subject for others, real harm may result, particularly when cumbersome and laborious methods are proposed in cases in which simple and direct formulae are already available.

A case in point is afforded by a recent paper by Reed.*

(1) The arithmetical routine indicated by Reed for the determination of the correlation coefficient by the summation of actual deviations, squares of deviations, and products of the deviations of the two variables from their respective means is wholly unsuited for practical work. All of the labor involved in the calculation of the five columns of his Table I is quite unnecessary. The elimination of this superfluous routine not only saves time but minimizes the danger of error which is particularly great when signs must be constantly regarded. The moments may be taken in terms of units about any point whatever. In these days of splendid calculating machines permitting simultaneous multiplication and summation, deviation methods involving positive and negative signs are inferior to one in which the first and second powers and the products of the two variables are summed, either directly from unclassified data or from the entries in a conventional correlation table. The formulae† required are:

$$\bar{x} = \Sigma(x)/N \qquad \bar{y} = \Sigma(y)/N$$

$$\sigma_x = \sqrt{\Sigma(x^2)/N - [\Sigma(x)/N]^2} \qquad \sigma_y = \sqrt{\Sigma(y^2)/N - [\Sigma(y)/N]^2}$$

$$r_{xy} = \frac{\Sigma(xy)/N - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

where x and y are the actual measures of the two variables, not their deviation from their mean value, the bars denote means and the sigmas standard deviations for the whole population of N individuals.

By the direct summation method Reed's Table I gives:

For July Rainfall, x, in inches:

$$\Sigma(x) = 243.9 \qquad \Sigma(x)/N = 4.065$$

$$\Sigma(x^2) = 1103.87 \qquad \Sigma(x^2)/N = 18.3978$$

$$\sigma_x = \sqrt{\Sigma(x^2)/N - [\Sigma(x)/N]^2} = 1.3687$$

*Reed, W. G. The Coefficient of Correlation. Quart. Pub. Amer. Stat. Asso., Vol. 15, pp. 670–684, 1917.

†Harris, J. Arthur. The Arithmetic of the Product Moment Method of Calculating the Coefficient of Correlation, Amer. Nat., Vol. 44, pp. 693–699, 1910.

Digitized by Google

For yield of corn in bushels, y,

$$\Sigma(y) = 2067.1 \qquad \Sigma(y)/N = 34.45$$
$$\Sigma(y^2) = 72446.69 \qquad \Sigma(y^2)/N = 1207.445$$
$$\sigma_y = 4.531$$

For rainfall and yield

$$\Sigma(xy) = 8612.99, \Sigma(xy)/N = 143.5498$$
$$r_{xy} = \frac{\Sigma(xy)/N - \bar{x}\bar{y}}{\sigma_x \sigma_y} = .565 \pm .059.$$

Reed gives $r = .526 \pm .063$, a value which is slightly erroneous because of the approximations necessary in the cumbersome method of calculation employed.

Full details concerning the application of this method to correlation tables, for both integral and graduated variates, are given in the paper cited.

(2) Reed lays entirely too great emphasis upon the necessity for a preliminary testing of linearity of regression and his discussion of the methods is very misleading. He says: "The correlations should never be attempted without first investigating the relationship far enough to see if it follows a straight line."

Now while it is true that the correlation coefficient is not strictly valid except in cases of linear regression it always gives at least a minimum measure of the relationship between two variables, and in the vast majority of cases this measure will be the one finally adopted after linearity has been tested.

Furthermore, the determination of the coefficient of correlation *is the first step in the critical testing of linearity.* It is quite unnecessary to plot the two variables in units of their standard deviation or to use the method of least squares in adjusting the line as suggested by Reed. When the correlation coefficient is known the straight lines are simply:

$$x = (\bar{x} - r_{xy}\frac{\sigma_x}{\sigma_y}\bar{y}) + r_{xy}\frac{\sigma_x}{\sigma_y}y,$$

$$y = (\bar{y} - r_{xy}\frac{\sigma_y}{\sigma_x}\bar{x}) + r_{xy}\frac{\sigma_y}{\sigma_x}x.$$

These can be calculated without even tabulating the raw data. Thus the equation from Reed's Table I is

$$y = 26.849 + 1.870\,x,$$

where y = yield in bushels and x = July rainfall in inches. This equation shows at once that the line in Reed's Diagram 2 on page 673 is incorrect. It should indicate an average yield of 30.59 bushels for 2 inches and 41.8 for 8 inches of July rainfall.

When the correlation coefficient and the regression equation have been determined, linearity may be tested by inspection of the distribution of the actual frequencies represented on a scatter diagram like Figure 2 or 3 of

Digitized by Google

Reed's paper, by a comparison of the empirical means of arrays with the theoretical means calculated from the regression equations, or by the use of Blakeman's criterion.*

Thus the calculation of the correlation coefficient, which by a practical method is a very easy task, should always be the first step in the analysis of the data. The nature of the regression line can then be determined in a convenient and really scientific manner.

(3) The bibliography given by Reed as "containing complete statements of the later development of work on the theory of correlation" is far from complete. A partial list of the omissions may be found elsewhere.†

This note is written with no desire to criticize the work of an individual writer, but merely in the hope of saving some beginner, who might follow Reed's suggestions, excessive and unnecessary labor in the calculation of the correlation coefficient, or misconceptions concerning the value or applicability of this most important statistical constant.

Cold Spring Harbor, N. Y. J. Arthur Harris.

COMMENT ON PROFESSOR IRVING FISHER'S ARTICLE ON THE "RATIO CHART."

The following remarks upon Professor Fisher's article in the *Quarterly Publications* for June are made without controversial intent and published rather as supplementary to than as critical of the statements made in the article in question. Where there is entire agreement, however, there is naturally little occasion for comment.

In the matter of terminology I do disagree with Professor Fisher, but that is hardly worth debating. The choice of the name "ratio chart," and the suggested opposition between "ratio" and "difference," is open to objection. Are not the curves in question currently and correctly known as logarithmic? Is it worth while to avoid this term simply because it, or the conception involved, is not familiar to the man in the street? All statisticians are acquainted with such curves, though many, it is true, feel them to be alien, just as some of them feel all graphic devices to be alien to their work. Should not the re-naming of a thing to make it popular be left to the commercial classes as a suitable mercantile, rather than a scientific or properly pedagogic, device? But, if the change means an improvement in our terminology, as regards either accuracy or adequacy, the case is different.

According to the usage with which I am familiar, statistical curves, and the diagrams constituted by them, are either arithmetic or logarithmic. Numbers and quantities are absolute or relative. Differences, also, may be absolute or relative. A relative difference is a species of ratio, and the

*Blakeman, J. On Tests for Linearity of Regression in Frequency Distributions. Biometrika, Vol. 4, pp. 332–350. 1905.

†Harris, J. Arthur. An Outline of Current Progress in the Theory of Correlation and Contingency. Amer. Nat., Vol. 50, pp. 53–64, 1916.

Digitized by Google

relative difference that is most familiar, both as regards species and form, is the per cent. increase. But the ratio may as well be to a constant as to a changing base—the latter being the case of per cent. increases. Index numbers are a familiar example of a system of relative numbers where the ratio is to a constant base. In a general way arithmetical curves and absolute numbers are thought of as corresponding to each other, the latter being the graphic expression of the former. This is not quite true, though there is no particular harm in such an association of ideas, provided it be accepted merely as such. Index numbers are relative numbers, but a series of such numbers, as directly plotted, constitutes an arithmetical curve. But a more thoroughly *relative* graphic expression of quantity and quantitative relations is found in logarithmic curves. I am informed by a mathematician to whom I have referred this paragraph that my use of terms here is in conformity with current mathematical usage.

The idea of the "ratio chart" as developed by Professor Fisher appears to correspond to that of relative numbers or rather, specifically, to per cent. increases. Such relative increases are so important in statistical analysis that a correlative graphic device properly claims an important place in statistical practice. It is true in a general way that logarithmic curves correspond to per cent. increases. The objection to an equation of the two is that per cents., though so serviceable, are fundamentally a clumsy and often misleading kind of relative number, while the logarithmic curve is a thorough and even elegant method of expressing quantitative relations where relative, rather than absolute, numbers are of fundamental significance—which is doubtless the more usual case in statistics. It is not a part of Professor Fisher's task to point out the limitations of per cent. comparisons, but the omission is significant if the reader gains the impression that the logarithmic curve is a translation of per-cent.-increase ratios into graphic terms. Relatively speaking, a 20 per cent. increase is not twice a 10 per cent. increase, and a 50 per cent. increase is not the equivalent of a 50 per cent. decrease. The defectiveness of the method of compiling index numbers formerly in use and recently abandoned by the United States Bureau of Labor Statistics was due to ignoring such facts. An increase of 20 per cent. in 10 years is not an average increase of 2 per cent. per year. These objections do not apply to the *logarithmic* expression of numerical relations. The peculiarity of the logarithmic curve—and of course the difficulty of understanding it—appears greater after considering these illustrative examples. But it should also be evident that this graphic and statistical device is much more worth mastering than it would be if it were merely the counterpart of numerical per cent. analysis. A true feeling for the significance of logarithmic curves is to be obtained only by studying and using them. But they are worth the trouble, and the limitations of per cent. analysis are the best proof of their worth.

Professor Fisher has passed rather too cursorily over a type of arithmetical curve in which the absolute numbers are not plotted directly, but instead per cent. indices. Per cent. indices are the quotients (with the decimal point moved two places to the right) of the absolute numbers,

Digitized by Google

resulting from division by some constant, the average being frequently the most convenient divisor. The writer is in the habit of calling such curves per-cent.-index curves. It should be noted that they use *relative* numbers. But the difference of such relative numbers from the absolute series is merely the difference between the foot and the meter—or whatever other arbitrary standard of length happens to be most convenient at the time being for purposes of plotting and comparison. The only difference between such a curve and a plot of the absolute numbers is in the scale, the preliminary computation serving chiefly the purpose of adjusting and controlling the scale, especially in a way to put two or more series of numbers upon an entirely comparable basis. Obviously such curves are to be interpreted like other arithmetical curves, though the adjustment of the scale is such as to make it usually convenient to refer the variations to a 100 per cent. axis instead of to a zero base, this feature justifying the cutting-off of the lower portion of the diagram and the utilization of the available space to better advantage, as may be done with logarithmic curves. Curves plotted from series of per-cent.-indices, it should be noted, are already adjusted so as to be grouped at the center of the diagram, at least if the 100 per cent. point for each series is the average of the series. Thus the per-cent.-index method has many of the advantages of the logarithmic method.

But the logarithmic curve is unquestionably a superior order of graphic device, and the fact that it is not generally familiar and understood is rather occasion for calling attention to its value—of course making the necessary explanation of how it works and why it is used—than for avoiding it where it is really most appropriate. That it is not always most appropriate may be illustrated by the fact that comparative logarithmic curves of the growth of population in France and Germany in the last hundred years would not meet the needs of the dominant interest in such a comparison. The per-cent.-index curve probably serves better to express the variation of an electrical load around its average than does a logarithmic curve, since the kilowatt capacity required at the peak should not, for most purposes, be represented as less important, or as occupying a less length on the vertical scale, than the kilowatt required during the hours of light loading.

That the use and qualities of logarithmic and per-cent.-index curves are subjects of thought not new to the writer may be shown by the following references to the annual reports of the New York Public Service Commission for the First District, where such curves, accompanied by more or less comment, will be found.* Logarithmic curves: Vol. III for 1908, p. 296ff; III for 1910, p. 24ff and p. 40ff;† III for 1911, p. 57ff; III for 1912, p. 78ff; II for 1913 p. 82ff.† Per-cent.-index curves: Vol. III for 1913, p. 71ff; III for 1914, p. 59ff and 72ff; II for 1915, p. 195ff.‡

New York City. G. P. WATKINS.

*From 1911 on the graphic work and a large share in the analysis is by Mr. L. H. Lubarsky.

†Here both the vertical and the horisontal scales are logarithmic.

‡This diagram has other special features, but the per-cent.-index device is fundamental.

Digitized by Google

NEW STATISTICAL BOOKS ADDED TO BOSTON PUBLIC LIBRARY.

Prepared by Library Committee: Roger W. Babson, Edmund E. Day, Henry J. Harris.

Quarter Ending March 31, 1917.

Bureau of Railway Economics, Washington. List of references on valuation of railways. August 1, 1916. Typewritten manuscript. [Washington. 1916.] Includes references on depreciation. *9385.a31

Employer's liability. Connecticut: Workmen's Compensation Commissioners. *9331.82a30

India. Department of Statistics. Statistical tables showing for each of the years 1901–02 to 1913–14 the estimated value of the imports and exports of India at the prices prevailing in 1899–1900 to 1901–02. Calcutta. 1915. *9382.54a7

Insurance almanac. An annual of insurance facts for 1916. New York. 1916. Tables. *9368.073a6

Jackman, William T. The development of transportation in modern England. Cambridge. 1916. 1 v. in 2. Maps. Tables. 9385.942

Jones, Grosvenor M. Ports of the United States. Report on terminal facilities, commerce, port charges, and administration at sixty-eight selected ports. Washington. 1916. Maps. Plans. [United States Department of Commerce.] *9382.73a36.33

Ricci, Umberto. Les bases théoriques de la statistique agricole international. [Traduit de l'original italien par M. de Brancion et M. Jacquemin.] Rome. 1914. Tables. 9338.1a24

Rubinow, Isaac Max. Standards of health insurance. New York. 1916. Tables. 9368.4a7

Smith, John Russell. Commerce and industry. New York. 1916. Illus. Maps. Plans. Tables. Deals largely with the United States. 9380.07a20

Times Book of Russia. Finance, commerce, industries. With an introduction by Sir Donald Mackenzie Wallace. London. [1916] Maps. 9314.7a24

Scherer, James Augustus Brown. Cotton as a world power. New York. [1916.] Tables. Authorities, pp. 426–436. Deals largely with the South. 9338.115a20

Underwriter Printing and Publishing Company. Record of business in each of the states and territories of the United States and the provinces of Canada by the joint stock fire insurance companies in 1915, and of aggregate business for thirty-five years from 1880 to 1914. New York. 1916. Tables. *9368.173a3

United States. Bureau of Labor Statistics. Retail prices, 1907 to December, 1914. Washington. 1915. Statistics of retail prices of food, coal and gas in the United States. 9331.073a24.156

Digitized by Google

Workmen's Compensation Act. Connecticut: Workmen's Compensation Commissioners. *9331.82a30

Quarter Ending June 30, 1917.

American Industrial Commission to France. Report to the American Manufacturers Export Association. September–October, 1916. [New York. 1917.] Illus. Portraits. Maps. "This Commission went to France to ascertain how American resources might best be made available for the reconstruction of the devastated regions." Bibliography, pp. 233–238. *9338.044a8

Chorlton, James Dewsbury. The rating of land values. Manchester. 1907. Tables. [Victoria University. Publications.] Deals principally with England and Wales. 9333.042a9

Grunsky, Carl Ewald, and Carl Ewald Grunsky, Jr. Valuation, depreciation and the rate-base. New York. 1917. Tables. Deals with public utilities. 9338.7a33

Frankel, Lee Kaufer. Some fundamental considerations in health insurance. [New York. 1916.] 9368.4a14

Juglar, Joseph Clément. A brief history of panics and their periodical occurrence in the United States. Third edition, translated and brought down to date, by De Courcy W. Thom. New York. 1916. Tables. 9332.75a14

Moore, John Monroe. The South today. New York. 1916. Portraits. Plates. Map. Tables. Deals with economic and social conditions in the South. 9317.5a2

North, Frank Allen. The income tax under United States and Massachusetts law. Boston. [1917] 9336.2473a14

Pickering, John Clark. Engineering analysis of a mining share. New York. 1917. 9338.2a7

Withers, Hartley. International finance. New York. 1916. Relates to Great Britain and to the United States. 9336.a4

SOME RECENT FRENCH STATISTICS ON INFANT MORTALITY.

At the outbreak of the war, effort looking toward protection of infant life in France was systematized by an official orgnization, namely, the Central Office of Maternal and Infantile Assistance. The purpose of this endeavor was "to secure to every pregnant woman and to every mother caring for a child under three years of age, suitable social, legal, and medical protection; to see to it that no woman should be forgotten and that no infant should remain unnoticed."

The effects of the numerous difficult war conditions on infant life were noted almost from the very beginning of hostilities. Dr. Adolphe Pinard, an obstetrician of the University of Paris, has condemned in toto the employment of pregnant women in industrial establishments, but his official confrères have not been able to adopt his extremely radical views. How

Digitized by Google

ever, the statistics which are responsible for these views are unimpeachable, although some of Pinard's colleagues of the *Academie de Medicine* allow themselves to be less alarmed than he. Pinard's figures, taken from the *Bulletin hebdomodaire de statistique municipale*, cover three years,—August 1, 1913–August 1, 1914, August 1, 1914–August 1, 1915, and August 1, 1915–August 1, 1916. Following is a brief review of the statistical material:

TOTAL NUMBER OF LIVE BIRTHS RECORDED AT PARIS.

August 1, 1913–August 1, 1914	48,917
August 1, 1914–August 1, 1915	37,085
August 1, 1915–August 1, 1916	26,179

Of the above the number of mothers concerned who receive public aid is thus indicated:

	Aid Given at Home.		Institutional Aid.	
	Number.	Per Cent.	Number.	Per Cent.
August 1, 1913–August 1, 1914	6,540	13.26	31,912	65.24
August 1, 1914–August 1, 1915	7,253	19.54	27,341	73.74
August 1, 1915–August 1, 1916	5,790	22.10	19,139	73.10

It will be noted that in the year August 1, 1915–August 1, 1916, 95.2 per cent. of all live births were assisted by public benefit. In the first year of the war, 2,491 women were delivered without public aid; in the second year, 1,250 women.

The tables quoted by Pinard for the three years mentioned fully discuss total live births, stillbirths, the number of infants handed over to nurses outside Paris, and infants abandoned by their mothers. Tables showing in detail infant mortality for 0 days—1 year, 1 year—2 years, and 0 days—2 years, and figures covering maternal puerperal mortality and causes, duration of pregnancy, and weight of newborn infants are also included in the paper; all of these refer respectively to the three years.

	Stillbirths. (Per Cent.)	Infants Placed Outside Paris. (Per Cent.)
August 1, 1913–August 1, 1914	7.69	31.00
August 1, 1914–August 1, 1915	6.63	13.35
August 1, 1915–August 1, 1916	7.57	22.32

	Infants Abandoned by Their Mothers.	Maternal Puerperal Deaths.
August 1, 1913–August 1, 1914	4.98	0.67
August 1, 1914–August 1, 1915	2.86	0.46
August 1, 1915–August 1, 1916	5.21	0.48

Digitized by Google

INFANT MORTALITY.

	0 days—3 mos. (Per Cent.)	0 days—1 yr. (Per Cent.)	1 year—2 yrs. (Per Cent.)	0 days—2 yrs. (Per Cent.)
August 1, 1913–August 1, 1914	8.10	15.51	4.33	19.86
August 1, 1914–August 1, 1915	7.41	15.20	5.12	20.36
August 1, 1915–August 1, 1916	7.40	15.00	5.50	20.48

Causes named are 10, namely, diarrhea and enteritis, congenital debility, respiratory affections, unknown or obscure diseases, tuberculosis, simple meningitis, measles, whooping-cough, scarlet fever, and diphtheria.

The result of Pinard's paper was to precipitate discussion in the Academy, considerable divergence of view being brought out. A committee was therefore appointed to study the subject and make suitable recommendations to the Academy. After debate on the measures advocated by a majority of the committee, the Academy modified the committee's resolutions which, somewhat amended, were finally adopted.

A committee of the Academy was then appointed to report on the subject of depopulation in France. A table included in this report, recently rendered, shows that the number of live births in France declined in France from 1,011,362, in 1876, to 745,539, in 1913. Some of the years enumerated show gain over the years immediately preceding, but not enough to offset the downward zigzag. Figures are also given for Germany, England, Sweden and Austria. The committee, representing primarily the standpoint of the physiologist, hygienist, and physician, concluded that the decline in the birth-rate can be remedied; that it is not due to impairment of French virility but that it is largely voluntary; that a child constitutes an expense and that the question is chiefly economic. Eight measures to relieve the situation are proposed by the committee; discussion by the Academy of these recommendations is pending.

REFERENCES.

Bull. de l'Acad. de Med., LXXVI, No. 50, p. 540.
Ibid. LXXVII, No. 1, p. 26.
" 5, p. 116.
" 7, p. 185.
" 8, p. 222.
" 9, p. 260.
" 10, p. 313.
" 11, p. 354.
" 12, p. 385.
" 20, p. 604.
" 22, p. 691.

Washington, D. C. THEODORE C. MERRILL.

Digitized by Google

Digitized by Google

AMERICAN
STATISTICAL ASSOCIATION

NEW SERIES, No. 120. DECEMBER, 1917.

POPULATION GROWTH IN SOUTHERN NEW ENGLAND, 1810–1860.

By Percy Wells Bidwell.

The most striking, indeed the essential fact in the economic history of southern New England in the half-century before the Civil War was an increase in population. Where, in 1810, 800,000 persons had with difficulty been able to get a living, 50 years later there were 1,900,000 on a much higher standard of life.

TABLE I.

POPULATION OF SOUTHERN NEW ENGLAND, 1810–1860.

Year.	Massachusetts.	Rhode Island.	Connecticut.	Total.	Density per Square Mile.
1810	472,000	77,000	262,000	811,000	58.3
1820	523,000	83,000	275,000	881,000	63.5
1830	610,000	97,000	298,000	1,005,000	72.3
1840	738,000	109,000	310,000	1,157,000	83.2
1850	994,500	147,500	371,000	1,513,000	108.8
1860	1,231,000	175,000	460,000	1,866,000	134.2

An analysis of population growth by decades from 1790 to 1860 reveals a distinct division of the 70 years into three periods: (1) from 1790 to 1820, (2) from 1820 to 1840, and (3) from 1840 to 1860.

TABLE II.

RATE OF INCREASE OF POPULATION BY DECADES, 1790–1860.

Decade.	Per Cent. Increase.			
	Massachusetts.	Rhode Island.	Connecticut.	Southern New England.
1790–1800	11.6	0.0	5.5	12.8
1800–1810	11.5	11.4	4.4	9.3
1810–1820	10.9	7.8	5.0	8.7
1820–1830	16.7	17.0	8.2	15.2
1830–1840	20.9	12.0	4.1	14.0
1840–1850	34.8	35.6	19.6	30.8
1850–1860	23.8	18.4	42.1	23.3

Digitized by Google

The steady tide of emigration which had been draining the towns of southern New England since the Revolution, and in some cases since 1750, was not perceptibly checked until after 1820.*

At the year 1840 we find a sharp turning-point; before this date, taking the three states as one region, the increase had been slow and uniform, averaging about 14 per cent. in a decade. After 1840 the rate of increase was nearly twice as great. The contrast between 1830–1840 and 1840–1850 is particularly striking. Of the population of Massachusetts a contemporary writer said:

> The fifty years since 1790, seem to be naturally divided into two periods of about equal length; the first extending from 1790 to about 1820, and the second from 1820–1840. During the first, the soil being nearly all occupied for the purposes of agriculture, the surplus population emigrated out of the state, and settled in other states. During this period, the increase was small, and the number of towns incorporated was comparatively small also. During the second period from 1820 to 1840, though the number of new towns was even less increased, the number of inhabitants was increased in a much larger proportion, than in the first. The surplus population, instead of emigrating, were more retained at home by the encouragements held out by the increase of the manufacturing enterprise of the community. The first period was distinguished by greater emigration to other states, and by the pursuit of agriculture as the principal employment, and the second by the increase of manufactures.†

The vital question for the economic historian when confronted by facts such as these is "How was the increase in population supported?" The agricultural industry in southern New England at the beginning of the nineteenth century had shown itself incapable of supporting in their accustomed manner of life the natural increase of people. Continuous emigration to western states was the result. But within a half-century not only had emigration ceased, but over 375,000

* For a discussion of the extent and causes of this early emigration, see the writer's Rural Economy in New England at the Beginning of the Nineteenth Century, Transactions of the Connecticut Academy of Arts and Sciences, Vol. XX, pp. 383–391, New Haven, 1916.

† Chickering, Jesse, Statistical View of the Population of Massachusetts, 1765–1840, Boston, 1846, pp. 41–42.

Digitized by Google

persons of foreign birth had been received.* By what change in the manner in which New England people got their living, by what modification of their "mores of self-maintenance" were the additional 1,000,000 persons in these three states supported? To answer this question is the purpose of the present investigation.

The Concentration of Population, 1810–1860. In the first place it is evident from a closer examination of the population statistics, that something occurred to cause a concentration of people in urban communities. Their growth accounts largely for the entire increase. In 1810 there were over the whole area of southern New England only three towns containing more than 10,000 persons.† Their total population was 56,000, amounting to 6.9 per cent. of that of the three states. About one quarter of all the people lived in 49 towns varying in size from 3,000 to 10,000, and the great remainder, two thirds of the total, lived in townships of less than 3,000 population. In the last group most of the people lived in farm houses scattered over an area of 40 or 50 square miles; only a very few lived in the villages.

The corresponding figures for 1860 show a marked contrast. At the latter date 26 towns of over 10,000 contained in all 681,600 persons, or 36.5 per cent. of the total population. Of the remainder one third was to be found in the towns from 3,000 to 10,000 and about an equal number in the towns of smaller size.

Looking at these changes from another point of view, it appears that most of the increase in population is to be ascribed to the growth of 26 towns of the largest size. To the total increase of 1,055,000 these contributed 625,600 or 59 per cent. The towns between 3,000 and 10,000 furnished 38.9 per cent., and the smallest towns about 2 per cent. The total population of the last group was only 19,000 greater in 1860 than in 1810.

* The number given by the Eighth Census, 1860, was 378,204. The statistics of internal migration published in the same census showed that between 1850 and 1860 the movement of the free native white population from state to state had resulted in a net gain to Connecticut of 17,500 persons; to Rhode Island, 3,500 persons. Massachusetts had a net loss of 6,650, thus leaving a net gain to southern New England of 14,350 persons. These figures are found in the volume on Population, p. xxxiii.

† Boston, Providence, and New Haven.

Digitized by Google

TABLE III.

CONCENTRATION OF POPULATION IN URBAN COMMUNITIES.

Groups.	1810.	1840.	1860.
Total population	811,000	1,157,000	1,866,000
Total number of towns	437	479	526
Group I.			
Towns over 10,000	3	9	26
Population	56,000	214,000	681,600
Percentage of total	6.9	18.5	36.5
Group II.			
Towns 3,000–10,000	49	74	129
Population	210,300	340,200	620,700
Percentage of total	25.9	29.4	33.8
Group III.			
Towns under 3,000	385	396	371
Population	544,700	602,730	563,700
Percentage of total	67.1	52.1	30.1

It will be noted that the middle-size towns remained a curiously constant factor in the total population. The increase in the importance of the largest towns was accompanied by a continuous decrease in the importance of the smallest.*

The causes of urban concentration are in general familiar: manufactures, the maritime industries, commerce, domestic and foreign, fishing, and shipbuilding. In southern New England hundreds of small towns and villages grew rapidly in population in the first half of the 50 years as a result of the establishment of manufacturing enterprises of all sorts. The growth of a few cities of 15,000 or more in eastern Massachusetts, such as Lowell, Lawrence, and Fall River, in the years 1840–1860, was a more spectacular exhibition of the same forces. The influence of the extension of maritime enterprises, on the other hand, acted somewhat differently. The effects on population in this case were naturally confined to a narrow strip of seacoast and localized at the best harbors. In only a very few cases can any direct effect of commercial expansion in increasing population be found. It is safe to say that unless manufactures had developed there would have been a stationary condition if not a decrease in the population of most of the commercial towns.

The Maritime Industries in their Relation to the Growth of Population. In tracing briefly the history of the maritime

* Identical towns were, of course, not to be found in the same groups in 1810 and in 1860, except in rare instances. Many of the smallest towns grew into communities of considerable size, their places in the group being taken by the incorporation of new towns through the subdivision of some of the larger and older.

Digitized by Google

industries of the ports of southern New England in the years 1810–1860, we may distinguish a number of separate branches: (1) foreign trade, (2) the coasting trade, (3) the cod and mackerel fisheries, (4) whaling, (5) shipbuilding and the manufacture of cordage. At the beginning of the century, every coast town from Newburyport along the shores of Massachusetts Bay to Cape Cod, and from New Bedford and Providence along the Sound to New York, where there was any sort of harbor, owned and manned a few small fishing and coasting vessels. Many of these towns traded with the West Indies; some of them, such as Nantucket, New Bedford, and New London, had built up a thriving whaling industry and a few, such as Salem and Boston, sent out a number of large vessels each year which came back freighted with rich cargoes from Europe and the Far East.*

The Foreign Trade. The foreign trade of New England ports suffered a severe decline in the early decades of the nineteenth century. The Embargo and Non-Intercourse acts of 1807 and 1808 brought to a sudden stop the lucrative carrying trade which our ships had been performing for belligerent European nations. Hardly had these restrictions been removed when the War of 1812 came on, causing a further interruption of trade until the cessation of hostilities in 1815.†

With the notable exception of Boston, almost every port in southern New England suffered severely; in many the loss of ships and of men was so great that the ports never recovered their former position. Salem had, in 1805, 41,600 tons of shipping; in 1816 only 34,000. Out of 152 vessels engaged in foreign trade owned in this port only 57 remained in 1815.‡

* See Rural Economy in New England, Chapter II, for a fuller discussion of the extent and character of this commerce.

† The extent of commercial loss appears from the following statistics taken from Macgregor, John, The Progress of America, 2 vols., London, 1847, II, 144, 198, 205.

	Registered Tonnage.			
Year.	Massachusetts.	Rhode Island.	Connecticut.	Total.
1806	306,100	28,600	26,000	360,700
1816	274,000	24,200	24,600	322,800
	Exports.			
	Massachusetts.	Rhode Island.	Connecticut.	Total.
1806	$21,200,000	$2,091,800	$1,716,000	$25,007,800
1816	10,100,000	612,800	594,000	11,306,800

The most notable decline was of course in the amount of foreign produce exported.

‡ Osgood and Batchelder, Historical Sketch of Salem, Salem, 1879, p. 134.

Digitized by Google

The tonnage registered in Newburyport shrank from 31,650 to 24,700; in Fairfield, Connecticut, from 6,200 to 2,500; and in New Haven from 11,600 to 6,700. Boston was the only port showing any substantial increase (from 115,200 in 1805 to 143,400 in 1816), although fishing ports such as Plymouth and Marblehead made slight gains.*

After a short period of rapid recovery, 1816–1825, the foreign commerce of southern New England ports showed little growth until 1840. In the two succeeding decades, however, the statistics of exports and imports and of registered tonnage show remarkable increases. Massachusetts merchants and shipowners controlled most of the foreign trade in this period. Rhode Island and Connecticut ports lost foreign trade steadily from 1830 to 1850, the latter state showing sharp recovery, however, in 1850–1860.†

TABLE IV. (a)
(000 omitted.)

States	1840.			1860.		
	Imports.	Exports.	Tonnage.	Imports.	Exports.	Tonnage.
Massachusetts	$16,500	$10,200	178.9	$41,200	$17,000	472.0
Rhode Island	275	207	21.5	496	221	10.8
Connecticut	277	518	9.0	1,420	743	20.4
Total	$17,052	$10,925	209.4	$42,116	$17,964	503.2

(a) From U. S. Commerce and Navigation Reports. The tonnage figures are registered tonnage, not including whaling vessels.

A closer examination of the statistics of tonnage registered in the various ports of Massachusetts in the years 1830 and 1860 shows that Boston not only absorbed all the increase in foreign trade but even drew away much of this trade from the smaller ports. Thus in Massachusetts in 1830 the registered tonnage of all ports, exclusive of whaling vessels, was about 180,000. Of this amount Boston had 107,000 tons. In 1860 of the total of 472,000 tons, Boston had 411,400. The increase for the state was 292,000 and for the port of Boston, 304,000

* The figures for 1805 are from Morse, Jedidiah, The American Gazetteer, Boston, 1810; and for 1816 from Spofford, Jeremiah, Gazetteer of Massachusetts, Newburyport, 1828; and from Pease and Niles, Gazetteer of Connecticut and Rhode Island, Hartford, 1819.

† The increase of the coasting trade and of whaling contributed somewhat to the growth of population in the Sound ports, especially in New London, Stonington, and New Haven. In the last, as well as in Providence and Bridgeport, the manufacturing industries seem to have been more important than commerce.

Digitized by Google

tons. Salem and Plymouth suffered most from this transfe in the former the loss was partly compensated by an increa in the coasting trade.

The Growth of Boston. It was this increase in commerci activity which was primarily responsible for population growt in Boston.

TABLE V.
INCREASE OF POPULATION IN BOSTON, 1810–1860. (a)

1810.	1820.	1830.	1840.	1850.	1860.
34,400	44,000	62,200	95,800	144,500	192,700

(a) Figures for the whole of Suffolk county, including the towns of Chelsea, North Chelsea, and W throp, as well as the city of Boston.

The period of most rapid increase in population, 1830–185 was also that of most rapid commercial expansion. Tl tonnage of this port engaged in the coasting trade doubled 1830–1840 and from 1840 to 1850 its foreign tonnage increase from 150,000 to 270,000. The first steam liner service betwee Boston and Liverpool was opened in 1840 and until 18 Boston was the exclusive American port of destination. few years later the famous Train Line of Liverpool packe was established. These two lines gave "an immediate ar very remarkable impulse to the general trade of the city, ar especially to its foreign business. . . . At no time ha there been so many importing and jobbing houses in Bost as between the years 1840 and 1857, although a few hous now do a larger business than was done by all of them the The coastwise shipping trade during this period was also ve large. The supplies of cotton for the mills of New Englar came by sea; also the products of the Southwest, such as pr visions, lard, lead, etc. In return we shipped to Southe ports domestic dry-goods, boots and shoes, furniture ar wooden ware; and when everything else failed there was alwa a tonnage supply of ice and granite for every vessel seeking cargo in that direction."*

The building of railroads in New England in the 20 yea before the Civil War was largely responsible for the conce tration of the commerce of the whole region in Boston. Mar

* From an article by H. A. Hill in Winsor's Memorial History of Boston, 4 vols., Boston, 1881, Vol. p. 228.

Digitized by Google

of the earliest roads had their termini in this city; others were soon connected with it. As early as 1845 we find Boston connected by rail with Lowell, Manchester, and with Nashua and Concord in New Hampshire; with Providence, Taunton, and New Bedford; with Marblehead, Salem, and Gloucester; with Portsmouth and Dover in New Hampshire; and with Portland, Maine; Springfield, Pittsfield, and Worcester, and with Albany and the Hudson River by the Western Railroad and the Boston and Worcester. At Worcester connection was made with the Norwich and Worcester road, built in 1840, and at Springfield with the road from New Haven, finished in 1844.*

In 1853 there were 39 lines of railroad completed in Massachusetts of a total length of 1,200 miles and 36 other lines under construction. A contemporary account reads: "By these roads Boston communicates directly with every important town in Massachusetts, and with most of those of the neighboring states. There are three lines of continuous railroads, uniting the capital of the state with New York and with the intermediate towns of Rhode Island, Connecticut, and central Massachusetts. Two other lines connect Boston with Portland and the towns between them. Two, one through Vermont, and the other through central New Hampshire, bring her into direct intercourse with Burlington, Vermont, and with Montreal, and with Ogdensburg; and another to Albany, opens an uninterrupted line of railway communication between Boston, Cincinnati, Terre Haute, and Chicago, and by the close of the present year (1853) will probably be united to St. Louis."†

Manufacturing enterprises in great variety, including iron manufactures of various sorts, machinery and steam-engines, stonecutting, leather dressing, painting, sugar-refining, and the manufacture of pianofortes had developed in Boston before 1860. According to the Massachusetts industrial census of 1855, these and other enterprises in Suffolk County employed 14,100 persons out of a total population of 172,000. Yet manufactures were relatively of less importance in Boston than in inland towns.‡

* See John Hayward's Gazetteer of Massachusetts, revised edition, Boston, 1849, pp. 422–434.

† From Lippincott, Grambo & Co.'s Gazetteer, Philadelphia, 1854, p. 676.

‡ This conclusion is based on a comparison of the proportions of the population engaged in manufacturing.

Digitized by Google

The New England Fisheries. New England fishe flourished in the half-century before the Civil War, furnish in their history the explanation of certain interesting popu tion changes. At the beginning of the nineteenth cent the cod and mackerel fisheries employed 1,200 vessels of 78, tons with crews of 10,360 men and boys, chiefly from the p of Maine (then a district in the State of Massachusetts) Massachusetts Bay.† In 1859 the number of vessels was 3,(measuring 157,000 tons with crews of 22,700. The indu was still divided between Maine and Massachusetts in ab equal proportions.‡

In Massachusetts up to about 1830 the principal fish ports were Gloucester, Beverly, Boston, Marblehead, N buryport, Plymouth, and the ports of Cape Cod. The hist of the next 30 years shows the constantly increasing importa of Gloucester, while the other ports, with the exceptio those on Cape Cod (Barnstable County), became comparati insignificant competitors. The total fishing tonnage of state was 61,700 in 1830 and 75,000 in 1860, the incr being 13,300 tons. In the same period the tonnage Gloucester increased by almost 27,000.

TABLE VI.

TONNAGE ENGAGED IN COD AND MACKEREL FISHING.

Ports.	1830.	1840.	1850.	186
Newburyport	7,360	4,871	6,553	6,
Ipswich	1,324	305	120	...
Gloucester	8,952	13,171	18,881	35,
Salem	2,709	5,583	169	...
Beverly	....	....	....	4,
Marblehead	5,362	4,617	2,936	5,
Boston	5,428	7,788	6,239	1,
Plymouth	8,047	7,380	6,631	4,
New Bedford	1,210	149	494	
Barnstable County	20,596	25,280	38,635	15,

in various counties in Massachusetts in 1855. Whereas in Middlesex County 10 out of every 57 v engaged, in Worcester, 10 out of every 46, and in Essex 10 in 30, in Suffolk County manufacturi ployed only 10 in every 122 persons. The difference is only partially accounted for by the inclusion statistics of the inland counties of some household manufactures which of course would not hav carried on in Boston.

† The fisheries of Rhode Island and Connecticut, with the exception of the whaling industry, were of importance. In 1807 all the fishing vessels of these states measured only 6,000 tons. In 1860 no vessels were registered in Rhode Island ports; in Connecticut the tonnage was 6,700. Macfarland mond, A History of the New England Fisheries, New York, 1911, p. 150, 153, 195–197.

‡ *Ibid.*, p. 174.

Digitized by Google

The influence of the fishing industry is clearly traceable in the growth of the population of Gloucester from less than 6,000 persons in 1810 to 14,141 in 1860.* Practically all of this increase must be ascribed to the extension of the maritime industry.† The history of the Gloucester fisheries shows great fluctuation in the several branches, but a steady progress as a whole. As soon as one branch proved unprofitable, another was developed, and when the fishing grounds in one region were exhausted, the enterprising skippers built larger vessels, extended their voyages and tried their luck in new waters.

From 1810 to 1820 there was little progress. After the interruption caused by the War of 1812, came a treaty more disastrous than the preceding hostilities, for by its terms the American fishermen were deprived of valuable treaty rights in Canadian and Newfoundland waters.‡ Moreover, the cod-fishing on the Grand Banks declined and consequently we find a small growth of population in this decade; 5,943 to 6,384. After 1820 shore cod-fishing§ became profitable and added to this was rapid growth in the mackerel catch. Scarcely had a decline become evident in this new branch when halibut began to be caught in great numbers on the Banks. When, just before 1840, the fisheries were seriously threatened by the departure of fish to more distant feeding grounds, the Gloucester men arose to the emergency, built larger vessels and extended their season. A few years later the off-shore cod-fishing revived.

In 1846 a railroad was completed connecting Gloucester with Boston. Cheaper and quicker transportation, together with the use of ice in packing fish for shipment, greatly extended the market for Gloucester fish and gave a new impetus to the industry. Just at the end of the 50's a new branch, the herring trade, was opened. In the early months of the year when the fishing fleet would otherwise be unemployed it began

* Including Rockport, originally a part of Gloucester but set off as a separate town in 1840.

† The only manufactures reported in the Massachusetts industrial census of 1855 were a cotton factory, employing 48 men and 135 women, and granite quarries employing 450 men. (Massachusetts, Secretary of State, Statistical Information Relating to Certain Branches of Industry, 1855.) The connection between fishing and granite quarrying is curious, the granite blocks originally taken out having been used as moorings for the fishing boats.

‡ Macfarland, New England Fisheries, pp. 154–155.

§ In 1832, 800 men were engaged in this branch alone. The Fisheries of Gloucester, 1623–1876, Gloucester, 1876, p. 30.

Digitized by Google

to make voyages to Newfoundland where large supplies of herring were purchased and transported to Boston to be sold partly for food and partly for bait.

The growth of population in Gloucester and Rockport was roughly parallel to the increase in the total tonnage of their fishing and coasting fleets, as appears from the following table:

TABLE VII.

POPULATION GROWTH AND TONNAGE STATISTICS, GLOUCESTER, 1810–1860.

Year	Population.	Tonnage.
1810	5,943	.....
1820	6,384	10,000
1830	7,510	12,600
1840	9,000	17,100
1850	11,060	22,500
1860	14,141	40,500

In 1855 over 80 per cent. of the entire male population between the ages of 15 and 60 were reported engaged in the fisheries.*

Barnstable County. This was always a unique region. Its infertile soil made an increase in agricultural population beset with even more difficulty than on the mainland. At the beginning of the century, a considerable portion of the food and fuel used by the inhabitants had to be imported from the mainland. Payment was made in the products of the fisheries and in the services of its coasting fleet. Without the development of any manufacturing enterprises,† the population of this county grew from 22,200 in 1810 to 36,000 in 1860. Here again an adequate explanation of population growth is to be found in the history of the maritime industries. The Cape Cod towns owned, in 1816, 16,000 tons of shipping, employed chiefly in cod-fishing. In 1851 the aggregate tonnage was 91,100, for the fishing fleet had more than doubled and, besides, a coasting fleet of almost 50,000 tons had been built up. The period of greatest increase in both these branches was 1830–1850; this was also the period of most rapid population

* This computation, including Rockport as a part of Gloucester, is based upon the Massachusetts Census of 1855. The proportion of males in the total population of Gloucester and Rockport was 51.6 per cent., which figure has been used in estimating the number of males in the above age-group.

† With the exception of a glass factory employing 500 persons and salt evaporation works employing 200 men. The latter industry had reached the height of its prosperity in the years 1835–1845.

Digitized by Google

growth.* After 1850 the aggregate tonnage fell off more than one third and population became practically stationary.

TABLE VIII.

BARNSTABLE COUNTY, POPULATION GROWTH AND TONNAGE STATISTICS, 1810–1860.

Year.	Population.	Total Tonnage.
1810	22,200	.....
1820	24,000	.....
1830	28,500	26,800
1840	32,500	56,600
1850	35,300	91,100
1860	36,600	63,600

Whaling Ports. The pursuit and capture of whales and the manufacture of whale oil and of spermaceti candles was a specialized branch of the fisheries. Larger ships, larger crews, longer voyages, and hence a much larger outlay of capital was necessary than in cod or in mackerel fishing. Consequently whaling, like all large-scale enterprises, tended to become concentrated at the best situations.† Nantucket, the leader until the War of 1812, lost half of its fleet of 46 vessels and its preëminence in that struggle.‡ Whaling, however, continued to be the principal occupation of the inhabitants of the island. In fact we find there the best example, with the possible exception of certain towns on Cape Cod, of a community dependent entirely on the maritime industries for its support. An official report in 1831 summarized the situation as follows: "The population of Nantucket is 7,225. For food, fuel, clothing and the material for building it depends on other places almost exclusively. Not a tree is fit for building; not a quarry; not a factory for cloth; not a piece of woodland, and but few cultivated acres, are to be found in an almost sterile island. Yet it is highly prosperous; but its resources are at an

* In 1850, the period of greatest prosperity for both the cod and the mackerel fleets, 8,865 men and boys, or considerably more than one representative from each family in the county, made up their crews. See Swift, Charles F., Cape Cod, An Historical Narrative, Yarmouth, 1897, pp. 315–316.

† There was also more distinct separation between owners and managers in whaling. The captain and the crew received wages, paid on a profit sharing plan, in part; the merchants and capitalists who owned and fitted out the ships remained on shore. The presence of a sort of leisure class in ports such as New London and New Bedford accounted for the cultivation there of the arts of life more than in other ports.

‡ Marvin, W. L., The American Merchant Marine, New York, 1902, p. 143 See also Morse, American Gazetteer, 1810.

Digitized by Google

immense distance in the Pacific ocean."* The climax of the prosperity of the island's unique industry came between the years 1840 and 1850; the decline which dates from this time was probably largely due to the increasing difficulty of fitting out vessels for long voyages at a port so far from the source of supplies. The lack of sufficient depth of water at its harbor entrance severely handicapped Nantucket in competition with New Bedford and New London.†

TABLE IX.

NANTUCKET, POPULATION AND TONNAGE OF WHALING FLEET, 1810–1860.

Year.	Population.	Tonnage.
1810	6,807	.....
1820	7,266	20,000
1830	7,202	19,817
1840	9,012	28,100
1850	8,452	19,050
1860	6,094	7,500

In 1860 the chief whaling port in New England, and in fact in the world, was New Bedford. From a fleet of 12 vessels which would not have averaged 250 tons each in 1805, this port had before the Civil War developed a whaling tonnage of nearly 130,000, employing 10,000 sailors. Whaling and the coasting trade, together with subsidiary industries explain almost entirely the increase in population in this port.

TABLE X.

NEW BEDFORD, POPULATION GROWTH AND TONNAGE STATISTICS, 1810–1860. (a)

Year.	Total Tonnage.	Whaling Tonnage.	Population.
1810	27,500	.....	5,651
1820	23,700	16,200	6,680
1830	51,400	25,900	10,626
1840	89,100	68,100	16,038
1850	131,400	96,400	20,747
1860	149,700	128,200	26,805

(a) The figures for 1810 and 1820, total tonnage, are for the nearest determinable date. For 1810 I have taken the statistics given by Morse in his American Gazetteer as of 1805 and for 1820 those given by Spofford, Gazetteer of Massachusetts, for 1818. The population figures for 1860 include the towns of Fairhaven and Acushnet, which were taken out of the original area of New Bedford.

* Documents Relative to the Manufactures in the United States. Executive Documents, 22 Cong., 1 Sess., Manufactures, Vol I, p. 202.

† Hunt's Merchants Magazine, XVII (1847), p. 377, mentions also a serious conflagration in 1846, destroying property valued at $1,000,000, which may have been a contributing cause.

Digitized by Google

Subsidiary industries such as the refining of oil and the manufacture of candles, the building of ships and equipping them with boats, masts, spars, sails, and cordage naturally grew up in such an important port.* Besides, cotton mills had been established in 1848 and enlarged in 1854, furnishing employment at that date for 500 persons.† It may be that the merchants and shipowners of New Bedford anticipated thus early the subsequent decline in the city's chief industry; at least it is true that among the names of the early financiers of the cotton manufacture are many which had long been associated with whaling.‡ The whaling industry as a whole had begun to decline just before 1860, due in part at least to overstocking the market with oil and bone. The rise in the price of these commodities had been continuous from 1830 until 1857, bringing steadily increasing profits to New Bedford capitalists. In the contemporary descriptions of the place we often find such statements as the following: "The whale fishery has proved very lucrative, and New Bedford is thought to be one of the richest cities in proportion to the number of its inhabitants any where to be found."§

The only other port of importance in whaling was New London. From very small beginnings, its fleet had increased to 77 vessels of 26,500 tons by 1845. Twenty-five of these ships in the next few years abandoned whaling for more lucrative trips to the California coast where gold had been discovered. Consequently the tonnage of the district in 1850 was only 11,500.‖ Some of the vessels returned to whaling, but the industry was never again important in New London. The growth of population which had been rapid in the years 1840–1850 (5,519 to 8,991) was considerably checked in the next decade (9,000 to 10,110). The decrease in whaling was to some extent offset by the establishment of a number of small manufacturing enterprises after 1852.¶

Shipbuilding and the Manufacture of Cordage. The shipbuilding industry reached the height of its prosperity in New

* Over 600 men were employed in these industries in 1855. Massachusetts Statistics of Industry, 1855.
† Crapo, William W., Historical Address, New Bedford, 1876, p. 51.
‡ Tower, Walter S., A History of the American Whale Fishery, Philadelphia, 1907, p. 75.
§ Hayward, John, Gazetteer of the United States, Boston, 1853, p. 471.
‖ Caulkins, Frances M., History of New London, 1612–1890, New London, 1895, pp. 644–645.
¶ See Adams, Sampson & Co.'s New England Business Directory, Boston, 1860.

Digitized by Google

England about 1850, when it furnished employment for 5,200 men in 366 shipyards. Between 1850 and 1860 the number of employees fell off almost one half and there was a corresponding decrease in the number of shipyards in operation. In 1860 the industry was almost equally divided between the states of Maine and Massachusetts, the latter having 89 shipyards with 1,400 employees. Medford, Newburyport, and Boston built practically all of the larger vessels in this state.* Only in the first of these towns can the influence of shipbuilding on population be clearly traced. Up to 1855, there was no other industry in Medford except shipbuilding and farming. Shipyards employing 160 men were established shortly before 1830 when the population of the town was only 1,755.† A small distillery employed three men, three hat shops employed 43 persons; the remainder of the population were farmers. In 1855 the shipyards employed over 1,000 men and the population had grown to 4,600. Practically no other industries had developed.‡

Accessory to shipbuilding was the manufacture of cordage. As early as 1810 Massachusetts was the leader in this industry in New England and in fact in the whole country. In 1860 practically all of the cordage of New England was made by 1,000 persons in 44 ropewalks in Massachusetts. The principal establishments were in Roxbury, Plymouth, and Boston, where their influence upon population growth was merged with that of manufactures in general, as well as with the other maritime industries.

Summary of the Influence of the Maritime Industries upon Population Growth. To summarize, it seems evident that in only a few coast towns was there considerably greater commercial or maritime activity in the years just before the Civil War than there had been in the years 1810–1830. Foreign trade had indeed increased, but its effects on population had been confined to the port of Boston. The fisheries had flourished with the extension of their market, but only the ports of Gloucester and of Cape Cod benefited thereby. The whaling

* Statistics of shipbuilding are found in the Censuses (U. S.) of 1850 and 1860 and in the U. S. Commerce and Navigation Reports for those years.

† Documents Relative to Manufactures, 1831, I, 346–347.

‡ A sash, door and blind factory employed 70 men. Massachusetts Statistics of Industry, 1855.

Digitized by Google

industry, too, was limited to New Bedford, Nantucket and New London. The coasting trade was the only branch of the maritime industries whose prosperity was widely dispersed. But even in this case, an increase in domestic commerce was often merely a compensation for the loss of foreign trade, and hence, as in the case of the ports on Narragansett Bay, an increase in population did not result from this cause. And finally, the increase in commercial activity of all sorts was to a large extent only the result of a more fundamental change, namely the growth of manufactures. A large part of the raw materials for the textile mills, the shoe and leather factories, the machine shops, and the metal manufactures came by water, either from other parts of this country or from abroad. Much of the food supply of the growing manufacturing towns, such as grain and flour, sugar and molasses, coffee, spices, etc., were transported by water to Boston and distributed thence by the railroads. That portion of the New England trade which was engaged in carrying the cotton of the southern states and some of the tropical products of the West Indies to Europe might have flourished had there been no industrial changes in the inland towns, but the entire increase in domestic commerce, and a considerable portion of foreign trade as well, was dependent upon the growth of manufactures.

The Rise of Manufactures and the Growth of an Industrial Population. England's industrial revolution of the last half of the eighteenth century was repeated in New England in the first half of the nineteenth. Economically the characteristic features were alike. On both sides of the Atlantic, the introduction of technical improvements, the increasing complexity of the division of labor, and the use of power-driven machinery brought about what is known as the factory system of production. The cost of transportation was lowered through the building of canals and railroads; with the market thus widened production was stimulated, old enterprises were enlarged, new factories established and the manufacturing population rapidly increased.

The social results following the establishment of the factory system of production in England were not as quickly apparent, nor in fact did they ever appear in such extreme form, in New

Digitized by Google

England. Our economic organization was much simpler in 1810 than that of the mother country had been in 1760. Most of our industrial products were turned out either in the farm-houses or in the shops of village artisans. Except in the shoemaking industry, the commission system had not been developed. Consequently there was in New England no industrial proletariat to be cruelly exploited. Hence we got our great industrial change without some of its more spectacular features.

Nevertheless, the growth of manufactures in southern New England in the years 1810–1860 was truly revolutionary. The establishment of mills and factories, on a small scale at first, in hundreds of small towns and villages and the subsequent growth of these enterprises were fundamentally responsible for a series of economic changes. It was the necessity for cheaper transportation facilities to bring raw materials to the factories, and to carry away their finished products, which led to the building of railroads connecting the industrial towns with the seacoast. It was the rise of an industrial population that must be fed which stimulated domestic commerce and gave a home market to the farmers. Until 1810 there had been practically no non-agricultural population in New England away from tide-water. Even the ten or twelve commercial towns contained but a small proportion of the total population. In these towns the nearest approach to manufacturing was found, but in them the industrial activities were mostly those of craftsmen, such as tanners and leather dressers, chandlers, coopers and cabinet-makers, shipcarpenters, etc. Such persons differed from the artisans of the country towns only in the fact that as their market was wider, so they were enabled to devote their entire time to a trade, thus becoming dependent on others for their food supplies. In the country towns, on the other hand, the artisans were regularly farmers as well, hence they cannot be classed as a non-agricultural, or food-consuming, element in the population. The beginnings of factory villages were to be found in only a few inland towns where such articles as tinware, clocks and shoes, for which a wide market existed, were produced by specialized workers.

Digitized by Google

The Causes of the Rise of Manufactures in New England. The causes of the rise of manufactures in the United States after 1810 are to be found in the circumstances which temporarily made it impossible to secure manufactured articles in any other way.* Before the interruption of foreign commerce in 1807, it had been the practice of the people in certain parts of the country to secure their manufactured articles from abroad. The tide-water plantations of the southern states were in the habit of exchanging their cotton, indigo, rice, and tobacco for the finer grades of clothing and house-furnishings. The Middle States and New England had no such great agricultural staples; for their grain, fodder and provisions there was no market except in the sugar plantations of the West Indies and in the rice plantations of South Carolina and Georgia.† The coast towns of New England were able to secure some foreign goods in exchange for the products of their fisheries and in payment for the services of their trading vessels in the carrying trade between Europe and the West Indies. For the farmers of the inland regions, both North and South, no such resources were available; whatever articles they found it impracticable to produce in the household, or in the shops of village artisans, they had to do without.

The interruption of commerce with Europe, in the years 1807–1816, suddenly cut off the supply of manufactures from abroad. The southern states could no longer export their staples; the northern skippers and fishermen found their industries paralyzed. It was this change in affairs which gave the embryo manufactures of New England their birth. The interruption of foreign commerce acted like a prohibitory protective tariff. It raised the prices of foreign-made goods to such a point as to make competition not only possible but extremely profitable. New England had the fundamental requisites for a manufacturing region. First of all it had an abundant supply of waterpower; it had also capital accumulated in commerce. When commerce was interrupted, this capital readily turned to investment in manufacturing.‡

* See Callender, G. S., Selections from the Economic History of the United States, Boston, 1909, Chapter IX, Introduction.

† See the author's Rural Economy in New England, pp. 294–304.

‡ A good example is to be found in the financing of the first cotton factories in Waltham and Lowell by

Digitized by Google

New England had the most concentrated as well as the most intelligent and energetic labor force to be found anywhere in the country. Its disadvantages, such as lack of technical knowledge and experience, the lack of improved machinery and the high cost of labor, were shared by other sections of the country which did not share the advantages.* The disadvantages, moreover, were but temporary in their nature; consequently the interruption of foreign commerce acted like a scientific application of protection to infant industries.

Early Stages in the Rise of Manufactures: 1810–1830. The new enterprises which sprang up with mushroom-like rapidity were but poorly equipped and were owned and managed for the most part by inexperienced men. Consequently the reappearance of foreign goods on the market after the War of 1812, in great abundance, at normal or less than normal prices, brought sudden disaster.† Many of the new enterprises were ruined, others suspended operations, and the whole upward movement in manufactures was checked. A period of readjustment followed, lasting until almost 1830. It was a time of elimination of the weakest concerns and of reorganization and rehabilitation of the stronger. After 1830 an era of real progress began, only temporarily interrupted by the crisis of 1837–1840, increasing steadily after 1840 and culminating in the brilliant decade, 1850–1860.

Progress in Manufactures, 1830–1860. In the years 1830–1860 there were two principal causes for progress in manufactures. In the first place there had been a great widening of the market through the increase of population in the newly settled regions across the Alleghenies, and through the increasing prosperity and purchasing power of the southern planters. The second general cause was the introduction of a number of important technical improvements, such as the power-loom in

Francis Cabot Lowell, Nathan Appleton and Patrick Tracy Jackson, merchants of Newburyport and Boston. See Cowley, Charles, History of Lowell, Revised edition, Boston, 1868, p. 37

* For a fuller discussion of these points, see Tudor, William, Letters on the Eastern States, New York, 1820, pp. 261–266.

† "Many of the manufacturers had rushed into the business, stimulated by the extraordinary rewards held out, with little or no knowledge of the methods and requirements of the manufacture, with insufficient capital, and with loose and unbusiness-like methods. All the circumstances warrant the belief that the contemporary records of the suffering and disaster which now befell them are not exaggerated." North, S. N. D., A Century of the American Wool Manufacture, in Bulletin of the National Association of Wool Manufacturers, XXIV, 234.

Digitized by Google

the textile industries, the Foudrinier machine in the paper industry, the use of sheet-brass in the manufacture of clocks, the Howe pin-making machine and sticking device, the use of the sewing machine in shoe-making and ready-made clothing industries, and the system of interchangeable parts in the manufacture of firearms, locomotives, steam-engines, and all sorts of machinery.*

Along with these improvements came a tendency toward production on a larger scale. Very few of the textile factories of 1830 employed as many as 100 hands; the great majority of them had less than 50.† A comparison of the figures given in the Censuses of 1840 and 1860 shows a decrease in the number of cotton mills in New England from 642 to 532 and an increase in the average number of employees from 60 to 122. Between the same dates the woolen mills decreased in number from 304 to 272 while the average number of employees in each increased from 28 to 82. The installation of new machinery and improved processes required the investment of more capital than many of the small concerns could command; hence they could not continue to compete with the newer and larger plants. When the market was still limited to the vicinity of each factory, advantages of situation were not decisive. But as the market widened those factories which were on a large waterpower, or whose managers could obtain a railroad connection, had such an advantage over their less fortunate or less foresighted competitors that they eventually absorbed most of the increase in business. After 1840 there is evident a distinct tendency for factories to concentrate in the towns which offered the best transportation facilities and in those in which a skilled labor force had collected, due to the long continued existence of a particular industry.‡

The Census of 1860 reported 20,700 manufacturing establishments in New England in which 392,000 persons were employed. Further analysis of the figures shows the bulk of

* The system of interchangeable parts, or the "uniformity system," was originated by Eli Whitney and first applied by him to the manufacture of firearms in New Haven, Conn., about 1800. An interesting description of the circumstances surrounding the development of the system is found in the Sketch of the Life of Eli Whitney, by William P. Blake, in the Papers of the New Haven Colony Historical Society, V, 116–126, New Haven, 1894.

† The statistics of manufactures collected in 1830 (Executive Documents, 22 Cong., 1 Sess.) show very clearly the small scale of manufacturing enterprises in New England towns.

‡ As, for instance, the britannia ware industry in Meriden, Conn.

Digitized by Google

manufacturing concentrated in the three southern states, Massachusetts, Rhode Island, and Connecticut. They had 12,400 establishments, employing 210,000 men and 103,000 women. The census officials were very liberal in their interpretation of the word "manufactures," including in the returns many occupations, such as basket-work, charcoal-burning, lumbering, blacksmithing, and the operation of saw-mills and grist-mills, all of which were in the rural communities but imperfectly separated from agriculture. In order to obtain a clear conception of the extent to which the growth of manufactures gave rise to a non-agricultural population, a strict interpretation of the word must be maintained, namely the business of producing goods for a wide market by the employment of specialized workers. Eliminating the pseudo-manufactures, we get the following figures showing the extent of manufacturing development in its principal branches in southern New England:

TABLE XI.

MANUFACTURES IN SOUTHERN NEW ENGLAND, 1860. (a)

Industries.	Number of Establishments.	Employees.
Cotton	530	65,200
Woolen	270	21,000
Other textiles	92	6,890
Boots and shoes	1,600	66,000
Ready-made clothing	390	14,300
Hats and caps	92	2,720
Paper	154	4,550
Printing and bookbinding	231	3,310
Iron; mining, casting, rolling, etc.	200	6,660
Brass and copper	178	5,700
Hardware and cutlery	384	13,100
Agricultural implements	129	2,000
Machinery	324	9,680
India rubber	16	1,200
Musical instruments	52	1,100
Cigars	105	1,250
Jewelry	148	3,200

(a) Taken from the Eighth (U. S.) Census, Manufactures.

Effects of Rise of Manufactures on Population Growth. The connection between the rise of manufactures and the growth of population may be shown first by a comparison of the population statistics in the various counties in southern New England between the years 1810–1860. All of the counties started on a fairly even basis as regards manufactures in 1810. By 1860 some counties had industries whose importance, as measured by the value of their annual product and by

Digitized by Google

the number of persons employed, was many times that of others. The manufacturing counties showed invariably the greater increase in population. In Connecticut, for instance, New Haven County had in 1860 manufacturing enterprises whose annual product was valued at $23,400,000, employing 18,800 persons; its population had increased 163 per cent. since 1810; in Litchfield County on the other hand, the annual value of the manufactures in 1860 was only $5,700,000; 4,500 persons were employed, and the increase in population in the half-century had been only 12 per cent. In Rhode Island a similar contrast is afforded by Providence and Washington counties. In the former the value of manufactures in 1860 was $29,200,000 and the number of employees 22,700; in the latter, manufactures were worth only $4,000,000, employing 2,700. The increase in population in Providence County had been 281 per cent. and in Washington County only 23 per cent. Massachusetts counties afford similar comparisons.

More valuable conclusions are obtained by limiting the scope of the inquiry to the smallest units of population, the towns. Taking first the group of 26 towns which in 1860 had a population of over 10,000 each, we find 12 containing 258,000 people, whose growth depended entirely on manufacturing; Worcester, Springfield, Lawrence, Lowell, Lynn, Taunton in Massachusetts; Providence, North Providence, and Smithfield in Rhode Island; Waterbury, Norwich, and Hartford in Connecticut.* In these towns 64,600 persons, or 25 per cent. of the population, were engaged in manufactures. A second division of the group is made up of 11 towns; Boston and its suburbs, Charlestown, Cambridge, Chelsea and Roxbury; Newburyport, Salem, and Fall River, all in Massachusetts; New London, Bridgeport and New Haven, in Connecticut. In all of these except Boston, the growth of population in the years 1840–1860 had been due almost entirely to the growth of manufactures; from 1810 to 1840, maritime industries had been more important. Of the 380,000 people in the 11 towns, 45,000, or 12 per cent., were engaged in manufacturing. Finally we find three towns, New Bedford and Gloucester, in Massachusetts, and Newport, Rhode Island, whose growth was due almost entirely to the prosperity of some branch of

*Providence, Norwich, and Hartford had before 1810 depended largely on commerce for their prosperity.

Digitized by Google

the maritime industries. Their population amounted to 43,700 in 1860 of which only 2,360, or 5.4 per cent., were in factories. Among the 23 towns whose growth was due, entirely or in large part, to the development of manufactures, many had already gained prominence in a particular line; thus Lowell, Lawrence, and Fall River and the Rhode Island towns were leaders in textiles, Providence in jewelry as well, Lynn in boots and shoes, Waterbury in the brass industry, New Haven in the manufacture of carriages and clocks, and Norwich in paper and textiles.*

The growth of Lowell is a striking illustration of the effect of manufacturing upon population. In 1820 there were on the present site of the city three villages, containing in all some 250 persons. The waterpower resources of the place had as yet attracted only a woolen mill and a few grist and saw mills. In 1822 the Merrimac Company secured control of the waterpower and began the building of a number of large cotton mills. Between that date and 1830, five large cotton manufacturing companies built mills at Lowell, as the place was now called, after Francis Cabot Lowell, one of the pioneers in its development. The population enumerated in the Census of 1830 was 6,474, of whom more than one half were factory operatives. The parallel growth of the number of factory employees and of total population is shown by the following table:†

TABLE XII.
POPULATION AND MANUFACTURING IN LOWELL, 1837–1860.

Year.	Total Population.	Factory Operatives.
1837	18,000	8,560
1845	25,160	9,235
1855	37,554	14,661
1860	36,827	13,206

* Statistics of the number of persons employed in manufactures in cities of 10,000 inhabitants and over are given in a letter from the Secretary of the Interior, Executive Document No. 29, 39 Cong., 1 Sess. (1866). The above conclusions are also based upon material found in the New England Business Directory (1860), in city directories and local histories, in Massachusetts Statistics of Industry, 1855; Statistics of Certain Branches of Industry in Connecticut, 1845; and Report upon the Census of Rhode Island, 1865.

† The facts in this paragraph are taken largely from Charles Cowley's History of Lowell. The population figures in Table XII for 1837 and 1845 are from Hayward's Gazetteer of Massachusetts (revised edition, 1849, p. 187); for 1855 from the Massachusetts State Census of that date, and for 1860 from the Eighth U. S. Census. The statistics of factory operatives for 1837, 1845 and 1855 are from the industrial statistics compiled in Massachusetts in those years and for 1860 from the Letter of the Secretary of the Interior, 39 Cong., 1 Sess., Executive Document No. 29, 1866.

Digitized by Google

Of the smaller towns in Group I (over 10,000), Waterbury, Connecticut, is a typical example. In 1820, when the brass industry first became permanently established there, the population of the town was 2,882, having been practically stationary since 1790. In 1860 the population was 10,004. The industrial growth of the town is outlined by a local historian as follows: "It was the metal button business that led to brass making, the industry upon which the town has chiefly been built up. . . . By 1840, the business of manufacturing sheet metal had taken the lead of all the others. Then began the fostering and development of other branches of business, the demands of which would consume this product. Pins, hooks and eyes, tubing, brass kettles, clocks, spoons and forks—the thousands of articles which can be made of sheet metal and wire, were added to the list of manufactures, as well as the incidental industries of machine making, acid making, casting, forging, and supplying the other things used and consumed in the various processes of the main production."* In 1845, the factory operatives numbered 727;† and in 1857, 2,730 were employed in five brass and copper mills and in 17 various sorts of subsidiary industries.‡

Out of the 129 towns in Group II (5,000 to 10,000), we may select for closer study the 47 whose population had grown to exceed 5,000 by 1860. Of these only three, Barnstable, Nantucket, and Brookline, all in Massachusetts, had failed to develop manufacturing enterprises supporting a considerable proportion of their inhabitants. Typical towns in this division were Danbury and Meriden in Connecticut. In the former the expansion of the manufacture of felt hats explains an increase of population from 3,600 to almost 9,000 (1810–1860); in this business more than one seventh of the inhabitants were employed at the latter date.§ Meriden had early been the seat of the tin-ware industry to which, in the years 1830–1860, the manufacture of britannia ware, silver-plated

* Anderson, Joseph, D.D., The Town and City of Waterbury, Ct., 3 vols., New Haven, 1896, II, 262–263.

† Statistics of Industry in Connecticut, 1845.

‡ Bronson, Henry, M.D., History of Waterbury, Waterbury, 1858, pp. 561–563.

§ Francis, W. H., History of the Hatting Trade in Danbury, Conn., Danbury, 1860, gives a valuable account of the origin and growth of this town's chief industry.

Digitized by Google

ware, ivory combs, cutlery, buttons, small iron articles, etc., was added. The population increase was from 1,250 to 7,500 in the half-century.* Among the towns showing most rapid growth were a few in eastern Massachusetts in which shoemaking employed a large proportion of the inhabitants.† Thus Randolph, in which over 1,500 persons were making shoes in 1855, showed a gain of 400 per cent. in population, 1,170 to 5,760; in Natick, 1,570 persons out of a total of 5,490 made shoes, the population growth from 1810 to 1860 being 610 per cent. Shoemaking was also largely responsible for the growth of Marlboro (1,675 to 5,900) and Woburn (1,220 to 6,290). The growth of 19 manufacturing towns, of this group, in Massachusetts was 312 per cent.; four towns in which maritime industries were more important, Barnstable, Beverly, Marblehead, and Plymouth, showed a growth of less than 40 per cent.

The dependence of population growth upon manufacturing can be shown also by taking certain typical counties and examining the history of each town separately. Thus it invariably appears that whatever increase in population took place in the county can be accounted for, at least in large part, by the growth of its manufacturing towns. In New Haven County, Connecticut, we find an increase in population of 60,300 in the half-century, of which more than one half, 32,300, is accounted for by the growth of the city of the same name. About 17,000 more was contributed by the growth of three manufacturing towns, Meriden, Waterbury, and Derby. Thus 49,100 out of the total increase of 60,300 is accounted for. The remaining 11,200 was distributed among a number of smaller towns, in practically all of which some manufacturing was carried on. In Windham County in the same state, the increase in the years 1810–1860 was 12,380, of which 11,000 can be accounted for by the growth of seven textile towns.‡

* In 1850, out of the total population of 3,560, about 600 were employed in the above-mentioned manufactures. See Davis, Charles H. S., History of Wallingford, including Meriden and Cheshire. Meriden, 1870, pp. 494–495.

† Shoemaking in these towns was not yet a factory industry; some of the processes, such as cutting the leather, were done in central shops, but the bulk of the work of sewing the uppers and attaching the soles was done by the country folk, men, women and children, either in their own homes or in small shops near their houses. See Hasard, Blanche E., The Boot and Shoe Industry in Massachusetts, in Quarterly Journal of Economics, XXVII, 236–262.

‡ Thompson, Killingly, Plainfield, Windham, Pomfret, Putnam, Scotland.

Digitized by Google

Applying the same method to several counties in Massachusetts, we get similar results. Thus in Plymouth County the gain in five decades was 29,600; to this increase five manufacturing towns contributed 23,500.* In Berkshire County the total increase was 19,300, of which 17,800 was supplied by the growth of six towns with flourishing manufactures.† In Bristol County, commerce as well as manufactures appears as a cause of growth. The growth of five towns accounts for 50,200 out of the total increase in population of 56,600.‡

Population Changes in Agricultural Towns. The foregoing pages have shown that wherever in southern New England, in the 50 years before 1860, there was an increase of population, the explanation may be found either in the expansion of maritime industries or in the establishment of manufacturing enterprises. But the converse of this statement is also true; wherever there was neither commerce nor fisheries nor factories, there, population was either stationary or declining. So widely were the small manufacturing businesses distributed before 1860 that in some counties such as Essex, in Massachusetts, and Windham, in Connecticut, not a single purely agricultural town could be found.§ In hilly inland counties, such as Franklin and Berkshire, in Massachusetts, on the other hand, the agricultural towns made up a considerable proportion of the entire number. A careful study of the occupations of the inhabitants of all the towns in southern New England reveals 112 in which agriculture was the sole industry; of these 68 were in Massachusetts, 38 in Connecticut, and 6 in Rhode Island. The movement of population in a typical group of Massachusetts towns (Franklin County) is shown in table XIII.

The declining tendency exhibited by this small group of agricultural towns, especially after 1830, is also shown by the entire group of 112 towns taken together.‖

* Abington, Bridgewater, Wareham, Hingham, and Plymouth.

† Adams, Great Barrington, Lee, Pittsfield, Stockbridge, and Williamstown.

‡ Attleborough, Easton, Fall River, Taunton, and New Bedford; the last being the only town in which the maritime industries were of great importance. For the growth of population in New Bedford see Table X, p. 825.

§ I have defined as "purely agricultural" towns with no manufactures for a wide market employing as many as 25 persons. The sources used in the selection of these towns were the industrial censuses of Massachusetts and Connecticut already cited, the gazetteers and local histories.

‖ The decline in the population of the purely agricultural towns does not necessarily indicate a decline

Digitized by Google

TABLE XIII.

POPULATION OF 14 AGRICULTURAL TOWNS IN FRANKLIN COUNTY, MASSACHUSETTS, 1810–1860.

	1810.	1820.	1830.	1840.	1850.	1860.
Ashfield	1,809	1,748	1,732	1,610	1,394	1,302
Bernardston	811	912	918	992	987	968
Charlemont	987	1,081	1,065	1,127	1,173	1,075
Gill	762	800	864	798	754	683
Colrain	2,016	1,961	1,877	1,971	1,785	1,798
Hawley	1,031	1,089	1,037	977	881	671
Heath	917	1,122	1,199	895	803	661
Leyden	1,009	974	796	632	716	606
Monroe	....	....	265	282	254	236
Rowe	839	851	716	703	659	619
Shutesbury	939	1,029	986	987	912	798
Sunderland	551	597	666	719	792	839
Warwick	1,227	1,256	1,150	1,071	1,021	932
Wendell	983	958	874	875	920	704
Total	13,881	14,378	14,145	13,639	13,001	11,883

TABLE XIV.

POPULATION OF 112 AGRICULTURAL TOWNS, 1810–1860.

	Massachusetts 68 Towns.	Connecticut 38 Towns.	Rhode Island 6 Towns.	Total 112 Towns.
1810	68,104	51,863	9,076	129,043
1820	70,910	51,207	9,618	131,735
1830	69,453	51,261	9,047	129,761
1840	67,771	48,601	7,463	123,835
1850	67,693	49,763	7,210	124,666
1860	65,577	50,229	7,373	123,179

in the total agricultural population, nor does it imply a decrease in the productivity of the agricultural industry in southern New England in the 50 years. It is possible that in many of the smaller manufacturing towns, such as those in Group II (3,000–10,000), the agricultural element may have increased in numbers. As for the productivity of agriculture, the natural supposition would be that, because of the stimulation of the home market, the farmers must have found larger production profitable. This increased production might have taken place, as it did in the North Central States after the Civil War, along with a decided decrease in the rural population. (See Hibbard, B. H., The Decline of the Rural Population, in QUARTERLY PUBLICATIONS OF THE AMERICAN STATISTICAL ASSOCIATION, New Series, No. 97, Vol. XIII, March 1912.)

Digitized by Google

WHOLESALE PRICES IN THE UNITED STATES, 1791–1801.

By Harold V. Roelse.

The aim of this work was to gain some definite idea of the trend of the price level following the adoption of the federal constitution. Mr. Alvin H. Hansen worked out indices of wholesale prices for the years 1801 to 1840, inclusive, in 1915,* and the original plan of the present work was to carry the study back to 1789.

To make the results comparable, Mr. Hansen's system was followed in general. The first step in the procedure was the attempt to secure monthly statements of wholesale prices of a considerable number of commodities. Yearly averages of these prices were then computed, and, using the year 1791 as a base, indices were computed for each commodity. A simple arithmetic average of the indices of all commodities was computed for each year, and these yearly index numbers were adjusted to the base 1860 to make them comparable with Mr. Hansen's curve.

The primary difficulty encountered was the scarcity of material. The only quotations of wholesale prices available for this work were those appearing in the Philadelphia *General Advertiser*, 1791–1793, inclusive, and in the Philadelphia *Gazette* and *Universal Daily Advertiser*, 1795–1801. No figures were found for the years prior to 1791, nor for the years 1794, 1796, and 1800. Reports were not available for twelve months of any year. The number of monthly reports and the number of commodities for which prices were secured appear in Table I.

In Table II are listed the commodities used and the yearly averages of the monthly quotations—in the order in which they appear in the *General Advertiser*. Some adjustments were necessary before these averages could be computed, such as adjusting changes in measures or quantities quoted. It was also necessary to change about half of the original figures

*Quarterly Publications of the American Statistical Association, December, 1915.

Digitized by Google

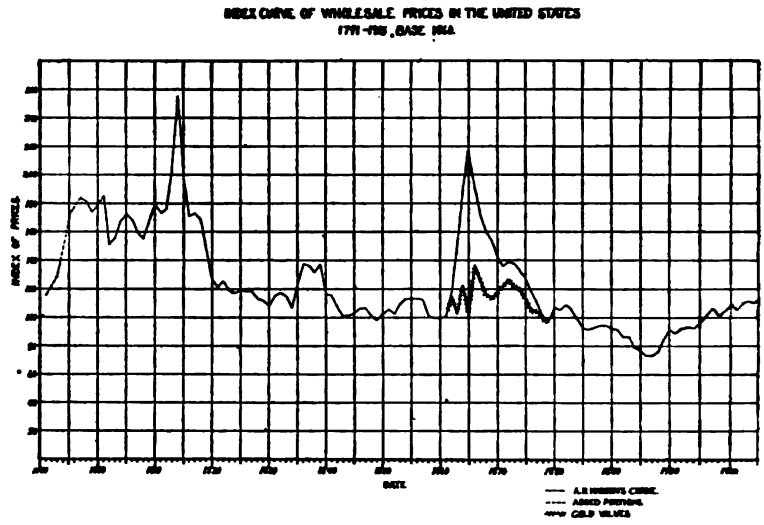
from the English monetary system to the present American system. The ratio used in this adjustment was seven and a half shillings to the dollar, which it appears was the ratio maintained in Pennsylvania during the period under consideration.

Table III represents the relative prices of these commodities for each year, index numbers having been computed using the year 1791 as the base. The average yearly index for all commodities appears at the end of the table and is taken to represent the relative level of prices for the period.

In order to join these figures to the curve plotted by Mr. Hansen it was necessary to find the ratio between them and his index numbers, adjusted to the base 1860 to make them comparable to the index numbers given in the Aldrich report. This was done in the following manner. First an index number was computed for Boston for the year 1801, as shown in the last columns of Tables II and III. Only one report could be secured for Philadelphia in that year—October thirty-first—so the Boston report was taken for about the same date. It will be seen that the average indices for the two cities are almost identical. While some prices were higher in Boston, others were correspondingly lower than in Philadelphia. The ratio of the Philadelphia index number for 1801 to Mr. Hansen's final index number for the same year was found to be 1 to 1.159. The average indices for the period 1791–1801 are therefore multiplied by 1.159.

Table IV contains the indices for the entire period 1791–1915, adjusted to the base 1860. Mr. Hansen's final indices are not tabulated so they had to be worked out from the original indices before the curve accompanying this work could be plotted accurately. His original indices were multiplied by 1.188 as he suggests. The index numbers of the U. S. Bureau of Statistics of Labor, base 1890–1899, were used for the period 1890–1913 in Mr. Hansen's work, and the present writer was able to add the indices subsequently published for 1914 and 1915. In adjusting these indices to the base 1860 they were multiplied by 1.098, as the Bureau of Labor reports for 1890 and 1891 are 1/1.098 of the indices for the same years in the Aldrich report.

Digitized by Google



INDEX CURVE OF WHOLESALE PRICES IN THE UNITED STATES 1791–1915, BASE 1860

The curve of wholesale prices represents the indices in Table IV graphically presented, and also the gold prices during the period of currency inflation from 1861 to 1879.

The existence of some inaccuracies in this work must be admitted. In the part of the curve that is original some slight errors were probably made in collecting the data. Toward the end of this period (1801), several varieties of some commodities, such as coffee, were quoted, while the earlier lists of prices contained very little in the way of specifications as to varieties or qualities of such commodities. An effort was made, however, to keep the same grade of each commodity throughout the work. Usually a high grade was chosen, but in the case of some commodities such as Madeira wine, a low grade was found to be quoted more steadily. The lack of complete data of course constitutes the chief weakness of this work. No high degree of accuracy is claimed for this portion of the curve, but it is believed that it represents approximately the course of prices for the period.

Where no material was available for a particular year, in plotting the curve it was necessary to connect directly the

Digitized by Google

points plotted for the preceding and following years. Referring to the curve, however, it will be noticed that in no case is more than one year at a time not definitely plotted, and that it is improbable that the price level actually deviated so widely for such years as to seriously affect the trend of prices during the decade. To be specific, in 1792 and 1793 the price level is evidently going upward quite rapidly; 1794 is missing, but in 1795 prices were very much higher and it is probable that in 1794 the prices were at least somewhat higher than in 1793; 1796 is missing, but following it are three consecutive years of high prices, so it is not probable that the price level was very low that year—it seems quite possible that it may have been even higher than the curve would indicate. As for the last missing year, 1800, a few reports for Charleston, South Carolina, show that prices were lower in that year than in 1801. The same was probably true in Philadelphia.

In conclusion it may be said that while this work can not be considered a highly accurate representation of the relative level of prices for specific years, it would seem to indicate conclusively that the decade following 1790 was a period of rapidly rising prices in the United States.

TABLE I.

Years.	Number of Monthly Reports.	Number of Commodities.
1791	11	68
1792	11	66
1793	8	68
1794	None	..
1795	10	58
1796	None	..
1797	5	59
1798	10	62
1799	3	56
1800	None	..
1801	1	56

Digitized by Google

TABLE II.

YEARLY AVERAGES OF WHOLESALE PRICES IN PHILADELPHIA (a), 1791-1801 (b).

Commodities.	Years.								
	1791.	1792.	1793.	1795.	1797.	1798.	1799.	1801.	1801. (Boston)
Ashes, pot—ton	$102.30	$100.00	$115.00	$141.33	$120.00	$124.80	....	$120.00	$140.00
Bread, ship-Pilot—cwt.	4.02	3.49	3.50	8.47	8.75	7.15	$8.50	....	8.00
Boards, Oak per M.	13.90	11.95	16.00	17.33	17.40	12.00	20.00	17.50	17.50
Boards, Merch. Pine—M.	15.23	19.30	23.00	25.33	25.33	24.00	26.00	13.00	14.25
Boards, Cedar—M.	14.19	16.73	18.00	20.00	20.00	24.00	26.00	....	....
Beef, Boston—bbl.	7.25	8.00	9.00	13.40	16.00	13.50	12.00	16.00	16.00
Beef, Country—bbl.	6.50	7.00	7.75	12.96	14.75	12.50	9.33	14.50	13.00
Butter—lb.	.149	.16	.16	.164	.15	.16	.16	.16	.26
Chocolate—lb.	.130	.176	.20	.26	.27	.25	.25	.25	.30
Coal—bu.	.215	.22	.225	.307	.375	.33	.303	.33	....
Coffee—lb.	.193	.19	.18	.224	.27	.25	.30	.23	.22
Cotton—lb.	.255	.378	.395	.31	.40	.40	.427	.40	.44
Copperas—cwt.	2.08	1.67	1.67	....	....	....	....	3.00	3.00
Cordage	8.28	8.00	8.25	....	....	....	....	16.00	16.00
Flour, superfine—bbl.	5.13	5.05	5.80	10.87	9.31	8.00	9.50	....	10.00
Flaxseed—hhd.	5.60	7.13	7.68	6.75	17.00	12.40	....	....	6.56
Glue—lb.	.155	.184	.20	.204	.17	.20	.18	....	.33
Gin, Holland—gal.	.767	.794	.885	1.27	1.50	.90	.97	1.30	1.12
Wheat—bu.	1.02	.944	1.16	1.93	....	1.34	1.70	1.67	....
Rye—bu.	.58	.512	.58	.955	....	1.00	....	1.00	1.00
Oats—bu.	.29	.278	.356	.41	.40	.40	....	.50	.55
Indian Corn—bu.	.41	.456	.547	.655	.73	.508	.49	.80	.83
Barley—bu.	.59	.888	1.02	1.09	1.27	1.33	....	1.00	1.12
Hams—lb.	.08	.09	.11	.122	1.45	.134	.11	.17	....
Lard—lb.	.09	.09	.10	....	....	....	....	.19	.19
Honey—gal.	.60	.60	.60	.90	.90	1.27	1.33	....	....
Hemp—ton	146.93	146.67	146.67	194.67	266.67	270.00	260.00	280.00	270.00
Hoops, hogshead—M.	15.67	15.00	15.00	30.00	29.33	29.00	29.00	16.00	20.00
Hides, raw—lb.	.11	.11	.11	.11	.103	.094	.10	....	....
Iron Bar—ton	80.50	83.06	86.25	85.60	114.50	98.33	98.33	130.00	120.00
Iron, Pig—ton	28.35	26.67	25.33	28.67	34.67	36.00	40.00	35.00	35.00
Iron, Sheet—ton	172.30	173.33	173.33	220.00	214.67	216.53	200.00	250.00	....
Herring—bbl.	2.75	2.85	3.00	....	6.00	5.60	3.83	7.50	4.25
Lead, Bar—cwt.	6.65	7.00	7.00	....	....	....	....	8.00	8.00
Leather, "soal"—lb.	.183	....	.20	.222	.20	.20	.20	.22	.21
Meal, Indian—bbl.	2.73	2.54	2.60	3.77	4.50	3.10	2.71	....	4.75
Mackerel—bbl.	8.27	8.82	8.00	11.58	12.75	12.00	9.00	11.00	9.50
Molasses—gal.	.37	.492	.51	.59	.70	.628	.647	.50	.49
Nails—lb.	.094	.10	.10	.122	.128	.13	.13	.12	....
Oil, Linseed—gal.	.50	.606	.654	.644	1. 9	.70	.657	.85	.92
Oil, Whale—gal.	.267	.267	.24	.40	.46	.47	.28	.61	.60
Oil, Olive—gal.	.86	.87	.87	....	....	....	....	1.87	....
Pitch—bbl.	2.32	2.00	2.00	3.01	4.67	3.50	3.00	5.00	5.00
Pork, Burlington—bbl.	11.67	11.09	12.94	16.33	18.75	18.20	15.00	26.00	26.00
Pork, Lower County—bbl.	10.45	10.91	11.56	15.07	16.00	15.50	14.00	24.00	23.00
Peas, Albany—bu.	1.09	.73	.73	1.00	1.13	1.00	.83	....	....
Pepper—lb.	.49	.422	.38	.40	.323	.298	.327	.29	.28
Pimento—lb.	.182	.17	.185	.16	.15	.17	.16	.16	.20
Rice—cwt.	2.88	2.71	3.11	5.15	3.58	2.73	2.33	6.00	6.25
Raisins—box	4.33	3.60	2.63	4.10	4.38	4.05	4.00	4.00	3.75
Rum, Jamaica—gal.	.83	1.135	1.12	1.16	1.60	1.40	1.22	1.40	1.20
Saltpetre—cwt.	12.00	13.75	14.33	....	....	42.50	50.00	30.00	30.00
Shot—ton	132.60	140.00	140.00	160.00	147.50	200.00	200.00	180.00	180.00
Steel, German—cwt.	9.29	9.18	9.00	9.20	10.92	11.00	14.33	....	....
Steel, American—ton.	126.27	113.21	113.33	133.33	127.33	156.00	156.00	81.00	....
Soap, Castile	.104	.11	.11	.28	.18	.20	....	.16	.16
Russia Sheeting	12.00	12.27	15.00	....	....	....	....	....	16.00
Sugar, Hav., white—cwt.	13.30	17.50	17.50	....	19.00	18.60	19.33	17.00	16.00
Sugar, Hav., brown—cwt.	10.00	15.00	11.75	13.33	16.50	13.60	14.67	13.00	12.50
Salt, Liverpool	.30	.306	.384	.70	.52	.68	.79	.58	.58
Shingles, Long—M.	10.22	13.00	13.00	20.00	20.00	18.00	18.00	....	....
Scantling, heart—M.	15.17	23.33	23.33	....	....	46.00	36.00	17.75	....
Staves, W. O. Hhd.—M.	19.25	19.42	20.33	37.87	44.25	35.00	35.67	16.00	30.00
Tar—32 gal. bbl.	1.66	1.50	1.50	3.20	3.33	3.25	2.20	4.00	4.50
Turpentine	2.50	2.00	2.165	3.21	4.00	3.50	2.50	3.00	5.00
Tobacco, James River	4.59	3.84	4.09	6.07	8.50	11.10	10.00	7.00	5.50
Tea, Hyson	1.12	....	1.28	2.00	1.70	1.44	1.53	1.13	1 10
Wine, Madeira—pipe	106.67	131.46	176.00	168.00	227.33	200.00	200.00	240.00	240.00

(a) The last column gives comparative prices for Boston.
(b) No figures were available for 1794, 1796, or 1800.

Digitized by Google

TABLE III.

INDEX NUMBERS OF YEARLY AVERAGES OF WHOLESALE PRICES IN PHILADELPHIA (a) 1791–1801, BASE 1791 (100).

Commodities.	Years.							
	1792.	1793.	1795.	1797.	1798.	1799.	1801.	1801. (Boston)
Ashes, pot—ton	97.75	112.41	138.16	117.30	122.00	....	117.30	136.85
Bread, ship-Pilot—cwt	86.82	87.08	210.80	217.65	177.88	211.46	....	199.00
Boards, Oak per M	85.98	115.11	12 .67	125.18	86.33	143.88	125.90	125.90
Boards, Merch. Pine—M	126.73	151.03	166.32	166.32	157.60	170.72	85.36	93.58
Boards, Cedar—M	117.90	126.87	140.96	140.96	169.14	183.25	....	....
Beef, Boston—bbl	110.35	124.14	184.84	220.70	186.23	164.53	220.70	220.70
Beef, Country—bbl	106.22	117.60	196.70	223.85	189.70	141.57	220.05	197.30
Butter—lb	107.40	107.40	110.07	100.07	107. 0	107.40	107.40	134.22
Chocolate—lb	135.38	153.85	200.00	207.70	192.31	192.31	192.31	153.85
Coal—bu	102.33	104.66	142.80	174.42	154.88	140.95	154.88	....
Coffee—lb	98.44	93.26	116.08	139.90	129.84	155.45	119.17	114.00
Cotton—lb	148.22	154.90	121.60	156.85	156.85	167.45	156.85	172.55
Copperas—cwt	80.30	80.30	....	....	....	....	144.23	144.23
Cordage	96.60	99.65	....	....	....	....	193.25	193.25
Flour, superfine—bbl	98.42	113.05	211.90	181.50	155.95	185.20	....	194.93
Flaxseed—bhd	127.32	137.13	120.54	303.55	221.42	....	....	117.15
Glue—lb	118.72	129.04	131.62	109.70	129.04	116.14	....	212.90
Gin, Holland—gal	103.54	115.40	165.60	195.60	117.35	126.50	169.50	140.04
Wheat—bu	92.56	113.75	189.23	....	131.40	166.70	163.74	....
Rye—bu	88.28	100.00	164.66	....	172.44	....	172.44	172.44
Oats—bu	95.87	122.78	141.40	137.95	137.95	....	172.42	189.68
Indian Corn—bu	111.22	133.42	159.78	178.05	123.90	119.58	195.12	202.45
Barley—bu	150.50	172.88	184.75	215.25	225.42	....	169.50	189.85
Hams—lb	112.51	137.52	152.53	181.30	167.53	137.52	212.54	....
Lard—lb	100.00	111.12	....	....	....	....	211.14	211.14
Honey—gal	100.00	100.00	150.00	150.00	211.70	221.70	....	....
Hemp—ton	99.83	99.83	132.50	181.50	183.76	176.97	190.56	183.76
Hoops, hogshead—M	95.74	95.74	191.46	187.18	185.07	185.07	102.10	127.64
Hide, raw—lb	100.00	100.00	100.00	93.65	85.46	90.92	....	....
Iron, Bar—ton	103.18	107.14	106.33	142.25	122.15	122.15	161.50	149.07
Iron, Pig—ton	94.08	89.35	101.13	122.30	127.00	141.10	123.45	123.45
Iron, Sheet—ton	100.61	100.61	127.70	124.60	125.68	116.08	145.11	....
Herring—bbl	103.63	109.10	....	218.18	203.72	139.26	272.73	154.54
Lead, Bar—cwt	105.26	105.26	....	....	....	....	120.30	120.30
Leather, "soal"—lb	....	109.30	121.32	109.30	109.30	109.30	121.32	114.75
Meal, Indian—bbl	93.04	95.24	138.10	164.84	113.56	99.26	....	174.00
Mackerel—bbl	106.65	96.74	140.03	154.18	145.10	108.84	134.22	114.88
Molasses—gal	133.00	137.85	159.43	189.20	169.74	174.88	135.15	132.45
Nails—lb	106.37	106.37	129.78	136.15	138.28	138.28	127.64	....
Oil, Linseed—gal	121.20	130.80	128.80	218.00	140.00	131.40	170.00	184.00
Oil, Whale—gal	100.00	89.88	149.82	172.30	176.05	104.90	228.50	224.73
Oil, Olive—gal	101.17	101.17	....	....	....	....	207.44	....
Pitch—bbl	86.20	86.20	129.74	201.30	150.86	129.32	215.52	215.52
Pork, Burlington—bbl	95.02	110.90	139.94	160.62	155.97	128.54	222.82	222.82
Pork, Lower County—bbl	104.40	110.63	144.22	153.12	148.33	134.00	229.68	220.10
Peas, Albany—bu	66.98	66.98	91.75	103.67	91.75	76.16	....	....
Pepper—lb	88.16	77.50	81.63	65.92	60.82	66.74	59.18	57.14
Pimento—lb	93.41	101.65	87.91	82.41	93.41	87.91	87.91	109.88
Rice—cwt	94.10	108.00	178.82	124.30	94.79	80.90	208.35	217.02
Raisins—box	83.14	60.74	94.69	101.15	93.53	92.37	92.37	86.60
Rum, Jamaica—gal	136.75	134.95	139.76	192.78	168.70	147.00	168.70	144.58
Saltpetre—cwt	114.60	119.42	....	....	354.20	416.70	250.00	250.00
Shot—ton	105.60	105.60	120.66	111.23	150.83	150.83	135.75	135.75
Steel, German—cwt	98.80	96.87	99.91	117.53	118.40	154.23	....	....
Steel, American—ton	89.65	89.75	89.75	100.85	123.55	123.55	64.15	....
Soap, Castile	106.73	106.73	269.23	173.10	192.31	....	153.85	153.85
Russia Sheeting	102.26	125.00	....	....	....	....	....	133.33
Sugar, Hav., white—cwt	131.58	131.58	....	142.86	139.85	145.33	127.82	120.30
Sugar, Hav., brown	150.00	117.50	133.30	165.00	136.00	146.70	130.00	125.00
Salt, Liverpool	102.00	128.00	233.33	173.30	226.67	263.30	193.33	193.33
Shingles, long—M	127.07	127.07	195.50	195.50	175.95	175.95	....	....
Scantling, heart—M	153.80	153.80	....	....	303.25	237.30	117.00	....
Staves, W. O. Hhd.—M	100.90	105.67	196.72	229.88	181.82	185.30	83.11	155.85
Tar—32 gal. bbl	90.36	90.36	192.77	200.60	195.78	132.53	240.98	271.10
Turpentine	80.00	86.60	128.40	160.00	140.00	100.00	120.00	200.00
Tobacco, James River	83.66	89.11	132.24	185.20	239.64	217.87	152.50	119.82
Tea, Hyson	....	114.30	178.51	151.80	128.57	136.61	100.90	98.20
Wine, Madeira—pipe	123.23	165.00	157.50	213.11	187.48	187.48	225.00	225.00
Average	105.62	111.74	147.72	158.69	156.47	150.12	159.33	159.54

(a) Last column gives comparative figures for Boston.

Digitized by Google

TABLE IV.

INDEX NUMBERS OF WHOLESALE PRICES IN THE UNITED STATES, 1791–1915, BASE YEAR 1860.

Date.	Index.	Date.	Index.	Date.	Index.	Date.	Index.	Date.	Index.
1791	115.9	1816	170.6	1841	115.8	1866	191.0	1891	92.2
1792	123.4	1817	172.7	1842	107.8	1867	173.2	1892	86.8
1793	129.5	1818	168.1	1843	101.5	1868	160.5	1893	86.8
1794	....	1819	148.1	1844	101.9	1869	153.5	1894	79.1
1795	171.2	1820	127.0	1845	102.8	1870	142.3	1895	76.9
1796	....	1821	121.5	1846	106.4	1871	136.0	1896	73.6
1797	183.9	1822	125.1	1847	106.5	1872	138.8	1897	73.6
1798	181.4	1823	118.9	1848	101.4	1873	137.5	1898	75.8
1799	174.0	1824	117.5	1849	98.7	1874	133.0	1899	83.5
1800	....	1825	118.8	1850	102.3	1875	127.6	1900	91.2
1801	184.7	1826	118.1	1851	105.9	1876	118.2	1901	89.0
1802	151.7	1827	118.3	1852	102.7	1877	110.9	1902	92.3
1803	155.9	1828	113.1	1853	109.1	1878	101.3	1903	93.4
1804	167.0	1829	111.8	1854	112.9	1879	96.6	1904	92.3
1805	172.1	1830	108.6	1855	113.1	1880	106.9	1905	95.6
1806	168.2	1831	114.9	1856	113.2	1881	105.7	1906	101.1
1807	159.6	1832	116.6	1857	112.5	1882	108.7	1907	106.6
1808	155.6	1833	115.0	1858	101.8	1883	106.0	1908	101.1
1809	....	1834	107.1	1859	100.2	1884	99.4	1909	104.4
1810	178.7	1835	123.3	1860	100.0	1885	93.0	1910	108.8
1811	173.1	1836	137.5	1861	100.6	1886	91.9	1911	105.5
1812	175.9	1837	136.3	1862	117.8	1887	92.6	1912	109.8
1813	204.8	1838	130.9	1863	148.6	1888	94.2	1913	111.0
1814	255.2	1839	136.9	1864	190.5	1889	94.2	1914	109.8
1815	200.4	1840	116.8	1865	216.8	1890	92.2	1915	112.0

Digitized by Google

THE CORRELATION OF HISTORICAL ECONOMIC VARIABLES AND THE MISUSE OF COEFFICIENTS IN THIS CONNECTION.

By Willford I. King.

He who reads statistical literature nowadays finds it literally teeming with studies in correlation. Apparently, the time has arrived when a statistician's ability is largely judged by his output of correlation coefficients. Unfortunately, however, many of the coefficients seem to be more valuable as mementoes of industry than for any new truth which they add to human knowledge. In fact, it may be said without exaggeration that only too often the statistician has been so entranced by the mathematical possibilities of the problem that he has lost sight of the real meaning of the operations involved and hence the conclusions presented have been utterly fallacious and entirely contrary to the facts. Is it not time to call a halt and ascertain just what ends may and may not be served by this newly popular mathematical aid?

It is apparently in the correlation of historical variables that statisticians have wandered astray most frequently. Economists have borrowed a new machine from mathematicians and have proceeded to use it without understanding its workings. As a result, strange and wonderful laws of economics have been deduced; figures have been made to lie most glibly, and all because the limitations of the device have not been recognized. Perhaps a brief explanation of correlation and its meaning may prove helpful to others who, like the present writer, may have searched long in the literature on the subject and found little enlightenment.

The first question is "What does the term correlation mean?" Evidently, it implies some relationship between things, and this particular relationship is always one of cause and effect. If a change in A causes a change in B, there is said to be correlation between them. Likewise if a variation in A causes both B and C to vary, there is said to be correlation between B and C. But, if B and C should, for entirely unconnected reasons, happen to vary in the same manner, there would be

Digitized by Google

no correlation between B and C, since the relationship is not a causal one.

One of the laws of physics is that a given force always produces identically the same effect whenever it is applied. Does this same rule hold good for economic or social phenomena? There is every reason to believe that it is exactly as valid here as in the other parts of the physical world. If this is not true, it is useless to hope for any material advancement of economic science. A law which only works part of the time is no law at all. Economics deals with physical beings and forces and it would be strange indeed if it should prove an exception to the general rules of the natural world. The chemist knows that 16 grams of oxygen and 2 grams of hydrogen will make exactly 18 grams of water. The physicist is certain that a force of one dyne will always give an acceleration of exactly one centimeter per second to one gram. They allow for no exceptions to the rule. The economist, on the other hand, has usually been satisfied to say that a lessening of the wheat supply by one tenth will increase the price, to an indefinite extent, perhaps 20 per cent., or perhaps 40 per cent. The natural scientist figures accurately; the economist ordinarily calculates but vaguely. Why is this true? The fact seems to be that the "natural" sciences are more highly developed and better understood than the science of economics. They are older and mathematics earlier lent them its aid. Economists are now really just beginning to utilize mathematics and their attempts in this line are still marked by the crudeness of the amateur.

One of the chief reasons why economists have put so little faith in the mathematical accuracy of the laws which they have formulated is that many of them have not fully comprehended the principle of the resultant of several forces acting simultaneously. They have failed utterly to appreciate or to utilize the concepts which are to the physicist the most elementary bases of the study of mechanics. They have attempted to deal with whole groups of variable forces at once instead of separating them carefully and studying the effects of each individual force by itself.

Isaac Newton long ago pointed out the fact that when several physical forces act on a body at the same time each

Digitized by Google

produces its own effect regardless of the others and that any motion arising is the resultant of all the independent forces compounded. Thus gravity, the bore of the gun, and the explosive force of the powder all act together in determining the movements of the projectile. The physicist has long ago separated these component forces and he can point out accurately the part played by each. Economists, on the other hand, are only beginning to realize that they must attack a problem in a manner identical with that used by the physicists.

To illustrate: If a steamboat which, in still water, would travel ten miles an hour proceeds upstream in a river flowing four miles per hour, the net result is but six miles an hour of actual progress. No physicist would think of attempting to measure the propulsive force of the engines by the distance traveled without first allowing for the effect of the river current. Economists have, however, habitually been guilty of making exactly this error. For example, they have attempted to correlate the cyclical movements of two variables, recording changes in business, without first removing the long time trends due perhaps to currency changes or population growth. They have tried to study the cycles of industry without recognizing the fact that they were dealing with a complex of distinct oscillations of different wave lengths and amplitudes rather than with one simple sine curve. They have asserted that, since the money curve does not correspond with the price curve, prices evidently do not vary exactly and proportionately with the money supply. All of these are clear instances of a failure to recognize the imperative necessity of excluding all forces but the one to be studied, before making any comparisons.

Another illustration may make this necessity more evident:

The change in wheat prices in 1916 was due to several independent forces, among which were:

1. Currency inflation.
2. The sinking of wheat in the Atlantic Ocean.
3. The commercial isolation of Russia.
4. Shortage of agricultural labor in Western Europe.
5. Poor climatic conditions for wheat.
6. Increased consumption in soldiers' rations.

Digitized by Google

A graph of the figures would show that wheat prices did not vary in the same proportion as the money supply. Neither did they vary inversely in the same ratio as the yield per acre. A study of the stocks of wheat on hand each month would show that the price did not conform to the theoretical supply and price curve worked out for this grain. The superficial critics would say that here was first class evidence of the inexactness of so-called economic laws. Yet it was doubtless true that each of the forces governing the price of wheat played its part with the same mathematical exactitude with which the hydrogen and oxygen united to form water. The price of wheat was a resultant of all the forces acting in unison. Why have economists been so slow to recognize the principle here illustrated?

One part of the explanation seems to be that it has been more difficult to separate for measurement the economic than the chemical or physical forces. In the economic field, it has not been feasible to control the conditions of the experiment. Because of this fact, it has been necessary to call upon the science of statistics to differentiate the effects of these separate forces. This new science has responded nobly, but, as yet, it has failed to accomplish the desired end with the precision attained by the chemist or physicist by experimental methods. Fortunately, however, this phase of the science is advancing so rapidly that the outlook for the future appears most encouraging.

If the assumptions just made are correct, if economic and physical phenomena are governed by laws equally exact, there can be no such thing as imperfect or partial correlation. Every correlation to exist must be perfect. Every cause must produce its effect exactly and invariably. If the cause and its effect are once completely isolated, the coefficient must always be unity. In every case, there either is or is not correlation between variable A and variable B—there can be no middle ground.

But it is common to obtain correlation coefficients of 0.3 or 0.6 or 0.7. What do these decimals mean? They merely indicate that the cause and its effect have not been completely isolated—that the effects of conflicting forces still enter in to

Digitized by Google

mar the results. The low coefficients, then, in no sense indicate imperfections in the actual correlation, but they do show conclusively that the statistician has failed in his efforts to exclude some of the untoward forces which ought to have been eliminated. This failure may have resulted from a non-comprehension of the complexity of the forces involved; from a lack of sufficient information to render the elimination of undesirable forces feasible; or from ignorance of the proper statistical methods which must necessarily be utilized in the process. Those troubled by errors arising from the cause last mentioned would be helped by reading the admirable articles in the June, 1917, issue of THE QUARTERLY PUBLICATIONS OF THE AMERICAN STATISTICAL ASSOCIATION, by Professors Irving Fisher and Warren M. Persons, in which they discuss the merits of some of the various methods of eliminating the effects of the forces which must be excluded before correlating the particular cause and effect under consideration.

The wave motions which are so universal in natural phenomena are just as common in economic fields. These waves have different lengths and amplitudes and the analysis of a compound of several such waves is no simple task. The economist will do well to study carefully the methods used in analyzing light and sound waves and the general characteristics of sine curves before attempting the simplification of the data at hand. He is almost sure to find that his problem is strikingly similar to those with which the physicists have grappled so successfully. Once the undesired forces have been completely eliminated, and the resulting figures have been plotted, two curves emerge—one representing the cause and the other the effect. If the elimination has been entirely successful, the fluctuations in the curves will correspond perfectly and the correlation will be evident at a glance. To compute a coefficient to prove whether correlation does or does not exist is, in this case, a manifest waste of energy.

In attempting by graphic methods, therefore, the correlation of two supposedly related historical variables, two curves are derived which either do or do not resemble each other closely. If they are very similar, all the correlation coefficients in the world will add nothing to the knowledge gained simply

Digitized by Google

by observing the curves. If they are dissimilar, it may be impossible to say whether there is or is not correlation. In this case, it is easy to tell in advance that a coefficient, if computed, will be low and will, therefore, prove nothing conclusively. In either instance, therefore, the coefficient has added nothing whatever to the knowledge obtainable from the graphic presentation. Too frequently, the coefficients presented are worse than useless for they lend the glamor of apparent erudition to a carelessly made study. Do coefficients, then, have any place in the correlation of historical variables?

So far as the present writer has been able to observe they have but one valid use. The size of the coefficient is merely a means of determining the closeness of fit of two curves. If there is a lag, that is if there is a time interval between cause and effect, it is frequently difficult to determine by the eye the exact lag which has actually occurred. In this case, by trying coefficients with lags of different lengths, it is possible to ascertain the approximate time interval which gives the best fit to the curves—in other words, the most common lapse of time required for the cause to produce the effect. Aside from this purpose, the labor of computing coefficients for historical economic variables seems to be largely wasted. When coefficients are used to cover up careless work in failing to eliminate the causes not under consideration, their use becomes inexcusable.

Strangely enough, many writers still seem to confuse correlation coefficients with the ratio of variation. It is entirely unnecessary for the ratio of variation to approach unity in order to give a high coefficient as the following simple examples will show.

RATIO OF VARIATION UNITY.

Size of Item.	Deviation from Arithmetic Average. x	Deviation Squared. x^2	Size of Item.	Deviation from Arithmetic Average. y	Deviation Squared. y^2	xy
32	−2	4	64	−4	16	+8
33	−1	1	66	−2	4	+2
34	0	0	68	0	0	0
35	+1	1	70	+2	4	+2
36	+2	4	72	+4	16	+8
		10			40	20

Digitized by Google

$$r_1 = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{20}{\sqrt{10 \times 40}} = \frac{20}{20} = 1.$$

RATIO OF VARIATION 0.5.

Size of Item.	Deviation from Arithmetic Average. x	Deviation Squared. x^2	Size of Item.	Deviation from Arithmetic Average. y	Deviation Squared. y^2	xy
32	−2	4	66	−2	4	+4
33	−1	1	67	−1	1	+1
34	0	0	68	0	0	0
35	+1	1	69	+1	1	+1
36	+2	4	70	+2	4	+4
		10			10	10

$$r_1 = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{10}{\sqrt{10 \times 10}} = \frac{10}{10} = 1.$$

In both instances the coefficients computed by Karl Pearson's method show perfect correlation, although the ratio of variation is entirely different. Further experimentation will demonstrate that changes in the ratio of variation have no effect whatever upon the correlation coefficients.

To sum up, therefore, it seems:

1. That it would be wiser for economists to go more slowly in applying coefficients of correlation to historical economic variables, and that such coefficients should be used mainly as an aid in measuring the length of a lag.

2. That no coefficient equals the graphic method for demonstrating whether correlation does or does not exist.

3. That more work needs to be done in devising statistical methods of eliminating all but one cause before comparing that particular cause with its supposed effect.

4. That the ratio of variation should never be confused with a coefficient of correlation.

Digitized by Google

THE CORRELATION BETWEEN A COMPONENT, AND BETWEEN THE SUM OF TWO OR MORE COMPONENTS, AND THE SUM OF THE REMAINING COMPONENTS OF A VARIABLE.

By J. Arthur Harris, *Carnegie Institution of Washington, Station for Experimental Evolution, Cold Spring Harbor, N. Y.*

Many quantitatively measured variables with which the statistician has to deal in both social and biological sciences are really composite in character. Thus the death rates of a series of districts are made up of mortalities due to a number of causes. The total milk production of a cow is the sum of the productions of individual lactation periods which may differ greatly in quantity. The total length of an organism is the sum of the lengths of its component parts. The annual egg production of a fowl in an egg laying competition is the sum of the records for individual months.

Thus $x_1, x_2, x_3 \ . \ . \ . \ x_n$ are the components of the variable X, where $X = \Sigma(x)$.

It is customary, and in many instances quite proper, to deal with all such constituent elements as quite independent variables. When only one correlation involving two individual components is to be determined, a table may be formed in the usual manner. Cases may, however, arise in statistical analysis in which it is desirable to determine the correlation between the magnitude of any individual constituent element, x, of the variable X and the sum of the remaining elements $(X-x)$. Indeed, all possible measures of this kind may be needed, *i. e.*, $r_{x_1(X-x_1)}, r_{x_2(X-x_2)}, r_{x_3(X-x_3)}, \ . \ . \ . \ r_{x_n(X-x_n)}$.

If X be large and the values of $x_1, x_2, x_3 \ . \ . \ . \ x_n$ variable, the arithmetical routine may be troublesome, in fact practically prohibitive.

The correlation between the sum of two components and the sum of the remaining components

Digitized by Google

$$\cdot \cdot \cdot \quad r_{(x_1+x_n)(X-x_1-x_n)},$$

$$\cdot \quad r_{(x_{n-1}+x_n)(X-x_{n-1}-x_n)},$$

ıore components and the sum
g.,

$$+x_4)(X-x_2-x_3-x_4), \quad \cdot \cdot \cdot$$

$$1-x_n),$$

permutations which increase
mponents, may be quite as
ation between any individual
emaining components.
ch relationships is relatively
ns between the components
ments have been determined.
suggest themselves to those
ble modifications of the corre-
vhile to put the equations on

ıs and standard deviations of
ents and of the difference be-
of two or more components.

d constant throughout. The
$\cdot \cdot \quad \sigma_{x_n}$ and r_{x_1X}, r_{x_2X}, r_{x_3X}
rmined. The standard devia-
nponent, say the pth, is

$$x_p - 2r_{x_pX}\sigma_X\sigma_{x_p}$$

of components the standard

$$r_{x_n}^2 + 2r_{x_1x_2}\sigma_{x_1}\sigma_{x_2} + 2r_{x_1x_3}\sigma_{x_1}\sigma_{x_3}$$

$$\sigma_{x_2}\sigma_{x_3} + \quad \cdot \cdot \cdot$$

nost convenient to work from
ıs the moments $\Sigma(x_1)$, $\Sigma(x_1^2)$,

437–438, 1909, and G. U. Yule, Introduction to the

t moment method of calculating the coefficient of
3–699, 1910.

Digitized by Google

$\Sigma(x_2)$, $\Sigma(x_2^2)$, $\Sigma(x_3)$, $\Sigma(x_3^2)$ . . . $\Sigma(x_n)$, $\Sigma(x_n^2)$ and $\Sigma(X)$, $\Sigma(X^2)$ with the product moments $\Sigma(x_1X)$, $\Sigma(x_2X)$, $\Sigma(x_3X)$ . . . $\Sigma(x_nX)$ lead directly to the desired results.

In the following section I give the equations in terms of the original moments or moment coefficients.

The means and standard deviations for the individual components, say for example the pth component, are

$$\bar{x}_p = \Sigma(x_p)/N, \quad \sigma^2_{x_p} = \Sigma(x_p^2)/N - [\Sigma(x_p)/N]^2$$

The means of the sums of the remaining components are quite obviously

$$\overline{(X - x_1)} = [\Sigma(X) - \Sigma(x_1)]/N, \text{ etc.},$$

while the standard deviations are given by

$$\sigma^2_{(X-x_1)} = \left\{\Sigma(X^2) - 2\Sigma(x_1X) + \Sigma(x_1^2)\right\}/N - \overline{(X - x_1)}^2,$$

$$\sigma^2_{(X-x_2)} = \left\{\Sigma(X^2) - 2\Sigma(x_2X) + \Sigma(x_2^2)\right\}/N - \overline{(X - x_2)}^2, \text{ etc.}$$

The mean for the sum of two components is

$$\overline{(x_p + x_q)} = [\Sigma(x_p) + \Sigma(x_q)]/N,$$

and so on for any number of components.

For the sum of the remaining $(n-2)$, $(n-3)$ components the means are

$$\overline{(X - x_p - x_q)} = [\Sigma(X) - \Sigma(x_p) - \Sigma(x_q)]/N,$$

and similarly for reductions due to the removal of 3, 4 or more components.

For the sum of two components, $(x_p + x_q)$

$$\sigma^2 = [\Sigma(x_p^2) + 2\Sigma(x_px_q) + \Sigma(x_q^2)]/N - \overline{(x_p + x_q)}^2$$

For three components, $(x_m + x_p + x_q)$

$$\sigma^2 = [\Sigma(x_m^2) + \Sigma(x_p^2) + \Sigma(x_q^2) + 2\Sigma(x_mx_p) + 2\Sigma(x_mx_q) + 2\Sigma(x_px_q)]/N - (x_m + x_p + x_q)^2$$

For four components, $(x_h + x_m + x_p + x_q)$

$$\sigma^2 = [\Sigma(x_h^2) + \ . \ . \ . \ + \Sigma(x_q^2) + 2\Sigma(x_hx_m) + \ . \ . \ . \ + 2\Sigma(x_px_q)]/N - (x_h + \ . \ . \ . \ + x_q)^2,$$

and so on for higher numbers of components.

The values of the means of the $(n-2)$, $(n-3)$, $(n-4)$ remaining components have been indicated.

Digitized by Google

The standard deviation of the value which remains after deducting the values of two components x_p and x_q is

$$\sigma^2 = [\Sigma(X^2)+\Sigma(x_p^2)+\Sigma(x_q^2)-2\Sigma(x_pX)-2\Sigma(x_qX)+2\Sigma(x_px_q)]/N - \overline{(X-x_p-x_q)^2}$$

The value of $\sigma_{(X-x_m-x_p-x_q)}$ is given by

$$\sigma^2 = [\Sigma(X^2)+\Sigma(x_m^2)+\Sigma(x_p^2)+\Sigma(x_q^2)-2\Sigma(x_mX)- 2\Sigma(x_pX)-2\Sigma(x_qX)+2\Sigma(x_mx_p)+2\Sigma(x_mx_q)+ 2\Sigma(x_px_q)]/N-\overline{(X-x_m-x_p-x_q)^2}$$

The process for the value remaining after the renewal of four components is

$$\sigma^2 = [\Sigma(X^2)+\Sigma(x_h^2)+ \ . \ . \ . \ +\Sigma(x_q^2)-2\Sigma(x_hX)- \ . \ . \ . \ -2\Sigma(x_qX)+2\Sigma(x_hx_m)+ \ . \ . \ . \ +2\Sigma(x_px_q)]/N-\overline{(X-x_h- \ . \ . \ . \ -x_q)^2}.$$

I now turn to the correlations.

Consider first of all the simplest of the two problems, the determination of the correlation between a component and the sum of the remaining components of the variable.

The correlations between the variable and its constituent elements r_{x_1X}, r_{x_2X}, r_{x_3X}, . . . r_{x_nX} are often wanted for themselves, and in any instance are relatively easily determined. The regressions of X on its constituent elements, say on the pth component, is

$$r_{x_pX}\frac{\sigma_X}{\sigma_{x_p}}$$

Obviously the regression slope of $(X-x_p)$ on x_p is

$$r_{x_p(X-x_p)}\frac{\sigma_{(X-x_p)}}{\sigma_{x_p}} = r_{x_pX}\frac{\sigma_X}{\sigma_{x_p}} - 1,$$

or in terms of correlation

$$r_{x_p(X-x_p)} = r_{x_pX}\frac{\sigma_X}{\sigma_{(X-x_p)}} - \frac{\sigma_{x_p}}{\sigma_{(X-x_p)}}$$

I now turn to the formulae necessary for the determination of the correlation between the sum of any two components, say x_p and x_q, and the sum of the remaining components,

Digitized by Google

$(X-x_p-x_q)$, of any three components, say x_m, x_p and x_q, and the sum of the remaining components $(X-x_m-x_p-x_q)$, and between the sum of any four components, say x_h, x_m, x_p and x_q and the sum of the remaining components, $(X-x_h-x_m-x_p-x_q)$.

The product moments for the sum of two variables (x_p+x_q) and the sum of the remaining components $(X-x_p-x_q)$ is

$$\Sigma(x_pX)+\Sigma(x_qX)-2\Sigma(x_px_q)-\Sigma(x_p^2)-\Sigma(x_q^2)$$

For the sum of three components, $(x_m+x_p+x_q)$, and the sum of the remaining components, $(X-x_m-x_p-x_q)$, the product moment is

$$\Sigma(x_mX)+\Sigma(x_pX)+\Sigma(x_qX)-2\Sigma(x_mx_p)-2\Sigma(x_mx_q)-2\Sigma(x_px_q)$$
$$-\Sigma(x_m^2)-\Sigma(x_p^2)-\Sigma(x_q^2)$$

The product moment for four components $(x_h+x_m+x_p+x_q)$ and the sum of the remaining components is of course

$$\Sigma[(x_h+x_m+x_p+x_q)(X-x_h-x_m-x_p-x_q)]$$

which is easily thrown into the convenient form illustrated above for two and three components.

In certain cases the coefficient of correlation between a single component, say x_p, and the sum of the components remaining after the deduction of the pth and one or more other components is desired. Or, more generally, the coefficient of correlation between the sum of n components and the sum of the components remaining after the deduction of $n+1$, $n+2$, $n+3$ . . . components may be required.

The moments from which the means and standard deviations of the various components, sums of components and differences may be computed, have already been indicated. Thus the product moments only are required.

For the correlation between a single component, x_p, and the variable less two components, x_p and x_q, the product moment is

$$\Sigma(x_pX)-\Sigma(x_p^2)-\Sigma(x_px_q)$$

For the correlation between a component, x_p, and the variable less three components, $(X-x_p-x_q-x_m)$, the product moment is

$$\Sigma(x_pX)-\Sigma(x_p^2)-\Sigma(x_px_q)-\Sigma(x_px_m)$$

Digitized by Google

The formula for the product moment underlying the correlation between a component and the variable less the sum of a larger number of components is obvious.

For the sum of two components, (x_p+x_q), and the value of the variable less three components, $(X-x_p-x_q-x_m)$, the product moment is

$$\Sigma(x_pX)+\Sigma(x_qX)-\Sigma(x_p^2)-\Sigma(x_q^2)-2\Sigma(x_px_q)$$
$$-\Sigma(x_px_m)-\Sigma(x_qx_m)$$

Correlations beyond these for the sum of three components, $(x_p+x_q+x_m)$, and the value of the variable less that of four components, $(X-x_p-x_q-x_m-x_h)$, will be required very rarely indeed. The product moment for such a relationship is

$$\Sigma(x_pX)+\Sigma(x_qX)+\Sigma(x_mX)-\Sigma(x_p^2)-\Sigma(x_q^2)-\Sigma(x_m^2)-2\Sigma(x_px_q)$$
$$-2\Sigma(x_px_m)-2\Sigma(x_qx_m)-\Sigma(x_px_h)-\Sigma(x_qx_h)-\Sigma(x_mx_h)$$

Thus the formulae for the determination of all of the fundamental constants required in the calculation of the coefficient of correlation between the sum of from two to four components and the sum of the remaining components have been deduced.

Some of these results are known, others are such as anyone familiar with elementary statistical theory might write for himself. My object here has not been to express the relationships in the most elegant form but in that most convenient for practical work. The computer will note the great advantages of determining the moments and product moments for the individual components once for all. It is often just such simplification of method that greatly decreases the labor involved in the routine of the computing room.

Digitized by Google

PERSISTENCY OF DEPENDENCY—A STUDY IN SOCIAL CAUSATION.

By Maurice B. Hexter, *Superintendent, United Jewish Charities, Cincinnati, Ohio.*

Partaking of the generally prevalent attempts of sociologists at "objectivity," social workers have of late years turned their attention to the scientific measurement of social phenomena. Such descriptions of processes and functions were desired as might be verified and checked up by any investigator provided that he approached his task in a scientific spirit. It was recognized that by such winnowing of facts the truth might be found. This assumes a recognition of the necessity of eliminating the personal equation—"subjectivity." "This 'impersonal' view, has indeed, alone made science possible, and until recently has been considered the adequate and secure foundation of scientific method in all fields." It is only lately that we have begun to realize the true significance of scientific methods in the sphere of philanthropy. Keeping of strict records and efficient investigation of cases has been urged only from the standpoint of protection of charitable institutions from the impositions of unworthy applicants; such functions have been utilized for the better exposition of the scope and efficiency of the given agency. Such data, of late, however, have been considered invaluable material for scientific investigation and precise presentation.

The present investigation comprises a statistical analysis of the case records of the United Jewish Charities of Cincinnati. An analysis of the duration of dependency presumes a centralized agency, for then each case is carried until solution; it is not transferred for treatment to another or allied organization. Thus, the organization comprises all the functions for rehabilitation work. This all-inclusive coördination and correlation makes it possible precisely to denote exact duration of dependency. Transients have not been considered within the scope of this investigation since there is no reliable information in one single organization relative to the duration of a tran-

Digitized by Google

Digitized by Google

Digitized by Google

sient's dependency. Similarly, cognizance has only been taken of those cases wherein results were positively ascertainable.

The investigation was undertaken for three general purposes.

1. Is the duration of dependency increasing or decreasing?

2. Are relief agencies tending to pauperize their constituency, understanding by that term a "subjective condition in which a person prefers and persists in living on charity, loses his respect for self-dependence and has no ambition to obtain through his own efforts, a more comfortable life?" In other words, does the relief office tend to engender a feeling of what Marx terms "*verdammte Bedürfnisslosigkeit*"?

3. Assuming that duration of dependency is one of the many indices or coefficients of efficiency in relief work, is there a possibility of developing a measuring scale whereby the work of various cognate organizations may properly be compared?

From a certain point of view, duration of dependency may be considered as withdrawals from a life insurance corporation and statistically treated similarly.* The following table, then, is homologous to a table of withdrawals in life insurance and the annual withdrawal rates are therefore analogous to the rate of solution of a case of dependency or in other words the rate of duration.

A word of explanation will make clear the meaning of the above table. Of 2,466 cases, 1,174 were solved with a duration of less than one year. The rate of solution, then, is 47.8 per hundred cases. Further, of 1,292 cases exposed to risk for a period of one year 385 were solved with a duration of one year. The rate, therefore, is 30 per hundred cases.

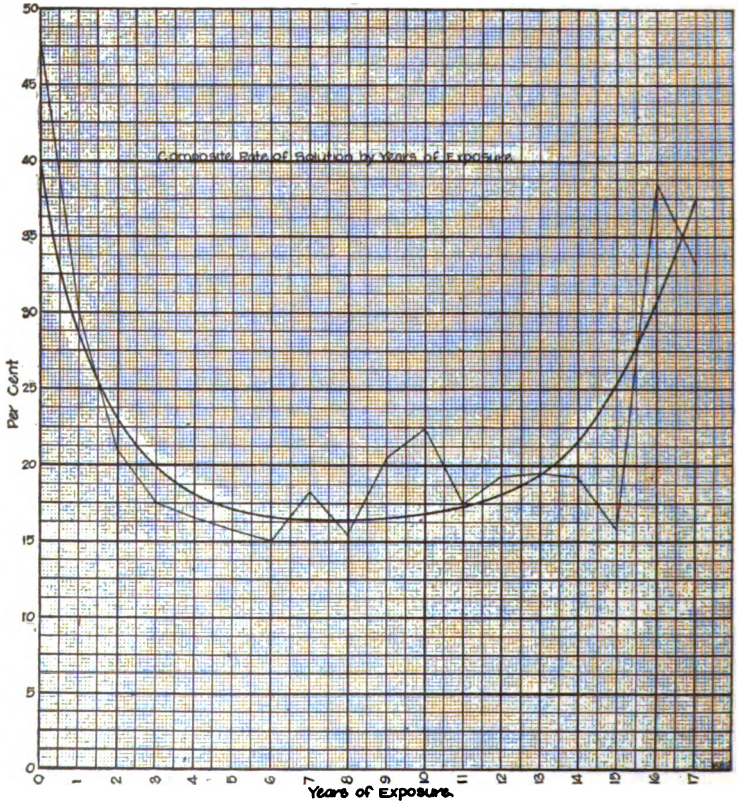
The light curve in the following chart is constructed from the actual rates of solution as given in Table III. The heavy is the smoothed curve through our rough data.† It can be seen that the rate of solution decreases rapidly for the first few years and then remains essentially the same for the fifth, sixth, seventh, eighth, ninth, tenth, eleventh, and twelfth years of duration. Thereafter the rise is rapid to the close of the seventeenth year of exposure. This may be ex-

* For a fuller explanation of the method herein used see Elderton and Pippard, Construction of Mortality and Sickness Tables, p. 29.

† For fuller explanation of the method see A. L. Bowley, Elements of Statistics, pp. 251–256.

Digitized by Google

plained by *a posteriori* reasoning somewhat as follows. The high rates of solution noticeable in the early years of exposure are due to the more or less acute cases of dependency. The medium rates of solution in the center of the mass of the data are due in large part to the chronic nature of the dependency, such as widowhood. The children are still small and no solution can be expected before the twelfth year of duration until two or three of the children have started to work. Color is lent to this hypothesis by the high rates of solution from the thirteenth year onward. Detailed examination of the cases for several years bears out directly this reasoning.



In addition to visualizing the rates of solution for our data, the curve may be used for comparative purposes. Thus, we

Digitized by Google

can determine whether any one single year gave a higher or a lower rate of solution than our average or composite. In this manner, we will be enabled to ascertain whether the duration of our dependency is increasing or decreasing. (To preclude unbiased errors we use the curve of actual average rather than the smoothed curve.) The following table shows the positive and negative deviations from the composite average. No account is taken of quantitative deviations but merely whether the deviation is positive or negative.

TABLE II.

POSITIVE AND NEGATIVE DEVIATIONS FROM THE COMPOSITE.

Year Case Opened.	Year of Exposure.																	Total Negative Deviations.	Total Positive Deviations.	Coincidences.	Total.
	0.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.				
1897–1898	−	−	−	+	−	+	−	+	−	−	−	−	−	−	−	−	−	0	14	3	1	18
1898–1899	+	−	−	−	+	−	−	0	−	−	−	−	−	+	−	+	+		11	5	1	17
1899–1900	−	0	−	−	−	+	−	−	−	−	−	−	+	−	+	+			11	4	1	16
1900–1901	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+				7	8		15
1901–1902	−	−	+	−	+	−	+	+	+	−	−	+	+	+					6	8		14
1902–1903	+	+	−	+	−	−	−	−	−	−	−	+	+						8	5		13
1903–1904	−	−	+	−	−	−	−	−	−	+	+	−							9	3		12
1904–1905	−	+	+	+	+	−	−	−	−	+	+								5	6		11
1905–1906	−	−	+	+	+	−	−	−	+	+									4	6		10
1906–1907	−	−	+	−	−	+	+	+	+										4	5		9
1907–1908	+	+	−	−	−	+	+	+											3	5		8
1908–1909	+	−	−	−	−	+	+												4	3		7
1909–1910	+	0	−	+	+	−													2	3	1	6
1910–1911	−	−	−	+	+														3	2		5
1911–1912	−	+	+	−															2	2		4
1912–1913	−	+	+																1	2		3
1913–1914	+	+																	0	2		2
Total																			94	72	4	170

In order to make clearer the data in the above table the deviations have been grouped into two parts: the first group comprises the series of years 1897–1898 through 1904–1905; the second group consists of the years 1905–1906 through 1912–1913. The following table shows this in concise form. From the table below it is noticeable that the proportion of negative deviations from our composite average decreases

Year.	Negative Deviations.		Positive Deviations.		Coincidences.	Total.
	Number.	Per Cent.	Number.	Per Cent.		
1897–1898 to 1894–1895	71	61.2	42	36.2	3	116
1905–1906 to 1912–1913	23	35.9	40	62.5	1	64

Digitized by Google

about 42 per cent. from the first period until the second period. In other words our rate of solutions of cases of various terms of exposure has materially increased in later years as is seen by the fact that of 116 observation periods 71, or 61.2 per cent., deviated negatively from our composite average, whereas in the second group there were 64 observation periods of which 23, or 35.9 per cent., deviated in a similar direction from our composite.

In all of the previous proofs of the decrease in duration of dependency we have formulated our comparison upon a composite average of all observations for all years and for varying terms of exposure. There remains another method of comparison, namely, to take as our basis the rates of solution of those cases arising during the year 1897–1898. In order to show the method involved the following table is shown. The structure of the table is such that we may ascertain what proportion of the cases originating in any one year have been solved after a definite term of exposure to solution. Thus, of the cases arising in the year 1900–1901, 71.3 per cent. have been solved after the fourth year of exposure, etc.

In order to simplify the table below we will tabulate the positive and negative deviations in a manner similar to the previous table where we tabulated the deviations from our

TABLE III.
CUMULATIVE RATES OF SOLUTION BY YEAR OF EXPOSURE.

Year Case Opened.	Year of Exposure.																	
	0.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1897–1898	43.7	54.1	62.4	69.4	73.6	78.6	80.0	84.2	85.6	88.4	89.8	90.5	91.2	92.6	94.0	94.7	96.1	97.4
1898–1899	53.9	59.0	63.1	69.2	74.3	75.3	77.3	81.4	83.4	86.5	87.5	89.5	90.5	92.5	93.5	94.5	97.5	
1899–1900	46.0	62.2	67.6	71.7	74.4	79.8	81.1	83.8	85.1	86.4	87.7	89.0	91.7	93.0	94.3	95.6		
1900–1901	51.4	58.0	62.0	66.0	71.3	80.5	84.5	85.8	88.0	90.6	91.9	93.2	94.5	95.8	97.1			
1901–1902	47.2	53.9	75.1	70.7	76.3	78.5	81.8	85.3	88.6	89.7	90.8	94.1	95.2	96.3				
1902–1903	49.0	70.2	75.7	83.7	84.9	86.1	87.3	88.5	89.1	90.3	92.2	94.1	96.0					
1903–1904	36.8	51.4	62.2	67.9	72.3	76.1	78.7	80.0	81.9	85.7	92.7	94.0						
1904–1905	41.2	61.7	73.1	78.6	83.6	85.5	87.4	88.3	89.2	92.4	94.7							
1905–1906	43.5	58.1	63.8	67.6	68.2	70.1	73.9	77.7	81.9	85.2								
1906–1907	42.8	55.0	66.1	69.5	73.8	78.2	83.2	87.6	92.0									
1907–1908	49.2	65.6	72.6	76.7	79.0	83.5	85.8	90.3										
1908–1909	52.2	63.9	67.4	71.5	75.0	80.9	87.3											
1909–1910	50.0	65.0	67.5	76.2	84.9	86.1												
1910–1911	40.3	53.3	59.1	66.6	75.2													
1911–1912	47.2	68.6	78.7	82.1														
1912–1913	40.3	61.9	73.4															
1913–1914	63.3	87.7																

Digitized by Google

compositive average. The following table shows the results of such a tabulation.

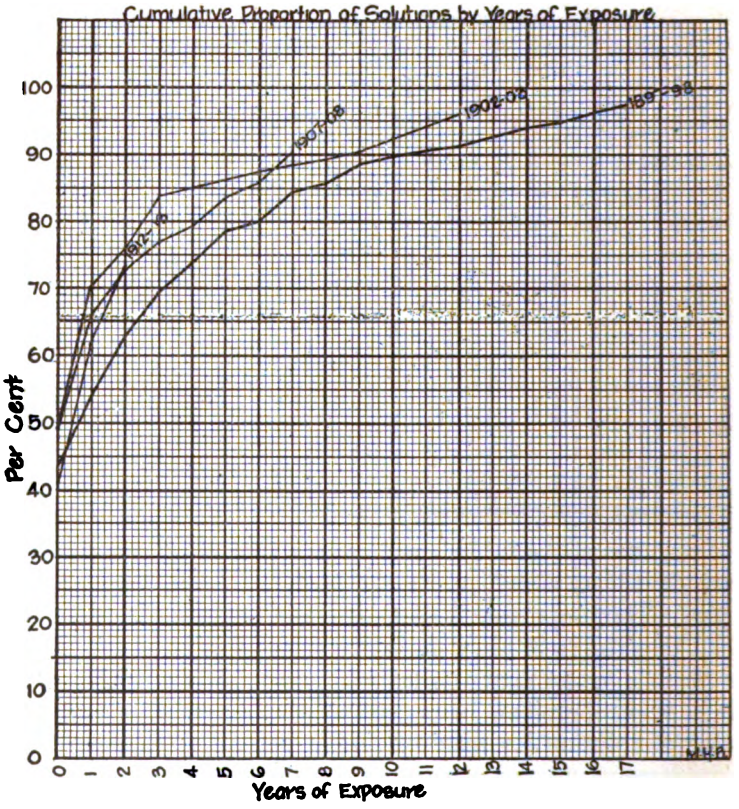
TABLE IV.
POSITIVE AND NEGATIVE DEVIATIONS FROM BASIC YEAR.

Year Case Opened.	Year of Exposure.																	Positive Deviations.	Negative Deviations.	Total.
	0.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.			
1898–1899	+	+	+	−	+	−	−	−	−	−	−	−	−	−	−	−	+		5	12	17
1899–1900	+	+	+	+	+	+	+	−	−	−	−	−	+	+	+	+			11	5	16
1900–1901	+	+	−	−	−	+	+	+	+	+	+	+	+	+	+				12	3	15
1901–1902	+	−	+	+	+	−	+	+	+	+	+	+	+	+					12	2	14
1902–1903	+	+	+	+	+	+	+	+	+	+	+	+	+						13	0	13
1903–1904	−	−	−	−	−	−	−	−	−	−	+	+							2	10	12
1904–1905	−	+	+	+	+	+	+	+	+	+	+								10	1	11
1905–1906	−	+	+	−	−	−	−	−	−	−									2	8	10
1906–1907	−	+	+	+	+	−	+	+	+										7	2	9
1907–1908	+	+	+	+	+	+	+	+											8	0	8
1908–1909	+	+	+	+	+	+	+												7	0	7
1909–1910	+	+	+	+	+	+													6	0	6
1910–1911	−	−	−	−	+	+													1	4	5
1911–1912	+	+	+	+															4	0	4
1912–1913	−	+	+																2	1	3
1913–1914	+	+																	2	0	2
Total																			104	48	152
Per cent.																			68.4	31.6	100.0

From the above table it will be seen that of 152 observation periods 104 deviated positively from our basic year—1897–1898. The proportion is 68.4 per cent. In other words, more than two thirds of the cumulative rates of solution are higher in periods subsequent to our basic year. Moreover the data we have compared are homogeneous, that is, we have compared rates of solution for cases of similar periods of exposure. It is interesting to notice that of the 108 observation periods in the first half of our data (namely from 1898–1899 to 1906–1907) 67, or about 62 per cent., deviate positively from the basic year. A similar calculation for the latter half of our data shows that of 44 observation periods 37, or about 84 per cent., deviate in a positive direction from our basic year. In other words, the rate of solution materially increases in the latter half of our data over the first half, taking the first year as our basic year for comparative purposes. This corroborates the comparison of our data with our composite average as shown in Table III. We see, then, that in all periods subsequent to our basic year the positive deviations far exceed the negative deviations. To restate the results, the rates of solution are

Digitized by Google

accelerated in later years, or in other words the duration of dependency is decreasing.



Summary.

To conclude, let us restate the threefold purpose of this statistical investigation. The first problem was to determine whether the duration of dependency is increasing or decreasing. The second problem was to determine whether relief agencies were fostering and engendering a state of pauperism. Thirdly, is there a possibility of developing a measuring scale of efficiency.

The answer to the first of these questions is easily given by the statement that the duration of dependency undoubtedly is decreasing. To assign causes for this decrease is difficult.

Digitized by Google

We may, I think, assume that the nature of the dependency with which we deal at present is at least as deep-seated and as chronic as that experienced fifteen years previous. Indeed, signs are not wanting that our current dependency is of a more chronic nature than hitherto. It seems that there is but one cause that will account for the constantly decreasing duration of dependency, especially since this decrease continues during years of cyclical fluctuations of labor. The cause, in our opinion, is to be found in the efficient rehabilitation that has been carried forward. We have set in action agencies which secure and discover cases of illness in their early incipiency. Thereafter, efficient coördination takes care of every emergency. On the side of illness, a convalescent home ends the rehabilitation. On the side of maladjustment to economic conditions a wholesome policy of establishment in business ventures has been prosecuted.

The second question can also be answered in the statement that relief agencies are not pauperizing their constituents as can be seen from the smoothed curve the shape of which can only be interpreted by the statement that those cases of long duration are more particularly, if not solely, composed of cases of long time dependency such as widowhood.

The third question relating to the possibility of developing a measuring scale or a norm is exceedingly theoretical. The first table presented in this study gave the rates of solution of cases of varying duration classed according to the year of origin. If similar studies were made over wide areas and by different organizations it would be possible to develop a table similar to the American Experience Table of Mortality in life insurance statistics. When such a table shall have been completed we can measure how far either above or below the expected rate of solution for any particular year of exposure a definite organization has reached. The corollary necessarily is that various organizations could be more intelligently compared than at present. It may be objected that the method is faulty, inasmuch as a large accession of widows with small children would tend to lengthen the duration of dependency. The answer to this objection is that the efficiency of an organization is to be found in the number of widowhoods that have been prevented.

Digitized by Google

REVIEWS AND NOTES

MINUTES OF THE FIRST MEETING OF PHILADELPHIA STATISTICIANS.

At the suggestion of Edwin W. Kopf, Assistant Statistician of the Metropolitan Life Insurance Company, and Carroll W. Doten, Secretary of the American Statistical Association, Dr. Chas. Scott Miller issued a call for a meeting of representative statisticians in the city of Philadelphia, together with local members of the American Statistical Association, to consider the advisability of organizing some statistical body in this city. Therefore, twenty-one invitations were extended and fourteen persons were present. The meeting was held on October 16 in the office of the Director of the Department of Public Health and Charities. Dr. Krusen, the Director, presided at the meeting and after expressing the purpose in issuing such a call, the subject was thrown open for general discussion. It was the concensus of opinion—

1. That there should be some standardization of method and improvement in statistical technique;
2. That data should be interpreted for public consumption in more of a non-technical manner than has been adopted heretofore.

It was moved by Dr. Jackson and seconded by Mr. Linton that a statistical organization be formed. The motion was carried. It was then moved and seconded that we affiliate ourselves with the American Statistical Association and establish a local branch of this body.

Temporary officers were chosen in the persons of Dr. Wilmer Krusen as Temporary Chairman and Dr. Chas. Scott Miller as Temporary Secretary. Mr. John A. Vogleson moved that the Chairman be authorized to appoint a committee of five on organization. Dr. Lee Welling Squier amended this motion by adding the Chairman and Secretary as members of this committee on organization *ex-officio*. The following were then appointed by the Chairman: Dr. Robert Riegel, Chairman; Prof. L. P. Myer, Dr. Louise Stevens Bryant.

The functions of said committee were defined as follows:

1. To draft a constitution and by-laws;
2. To plan for additional membership;
3. To suggest the number and method of conducting meetings, both for this body and for public gatherings;
4. To stimulate interest in statistical endeavors along all lines of work.

It was suggested that members send the names of any persons who might be interested in statistics to the Secretary and that an invitation be extended to these persons to attend our next meeting.

On motion of Dr. Francis D. Patterson, which was properly seconded, the meeting adjourned to meet at the call of the chair.

Digitized by Google

The following persons were present:

Dr. John Price Jackson, Commissioner, Pennsylvania Department of Labor and Industry;

Dr. Robert Riegel Assistant Professor of Insurance and Statistics, University of Pennsylvania;

Mr. George Wood;

Mr. George M. Merkle, Statistician, Division of Housing and Sanitation;

Miss Alice Haldemann, Bureau of Municipal Research;

Mr. William Byer, Bureau of Municipal Research;

Prof. L. P. Myer, Department of Economics, Temple University;

Dr. Louise Stevens Bryant, Municipal Court;

Dr. Lee Welling Squier;

Mr. M. Albert Linton, Vice-President, The Provident Life and Trust Co.;

Dr. Francis D. Patterson, Chief, Division of Hygiene and Engineering, Pennsylvania Department of Labor and Industry;

Dr. Wilmer Krusen, Director, Department of Public Health and Charities;

Mr. John A. Vogleson, Chief, Bureau of Health;

Dr. Chas. Scott Miller, Chief, Division of Vital Statistics.

CHAS. SCOTT MILLER, M. D.,
Secretary.

STATISTICAL BOOKS ADDED TO THE BOSTON PUBLIC LIBRARY DURING QUARTER ENDING SEPTEMBER 30, 1917.

PREPARED BY LIBRARY COMMITTEE, ROGER W. BABSON, EDMUND E. DAY, HENRY J. HARRIS.

Bancroft, Hugh. Inheritance taxes for investors. 2d edition. Boston. 1917. Tables. 9336.231a5

Barker, Harry. Public utility rates. New York. 1917. Tables. 9381.09

Bradbeer, William West. Confederate and Southern State currency. Mt. Vernon, N. Y. 1915. Facsimiles. 9332.5731

Brown, Harry Gunnison. Transportation rates and their regulation. New York. 1916. 9385.973a132

Canada. Department of Trade and Commerce. Reprint of articles dealing with Russian trade. Preliminary report. Ottawa. 1916. Maps. *9382.71a12

Chapman, Sydney John, and Harry Mainwaring Hallsworth. Unemployment: the results of an investigation made in Lancashire. Manchester. 1909. Tables. [Victoria University. Publications.] 9331.9042a2

Drachmann, Povl. The industrial development and commercial policies of the three Scandinavian countries. Edited by Harold Westergaard. Oxford. 1915. Tables. [Carnegie Endowment for International Peace. Division of Economics and History.] 9338.048

Digitized by Google

Guinchard, J., editor. Sweden. Historical and statistical handbook. 2d edition. Stockholm. 1914. 2v. Illus. Portraits. Maps. Facsimiles. *9314.85a48

Guenther, Louis. Investment and speculation. Chicago. 1916. Plates. Charts. Tables. [La Salle Extension University. Business administration text-books.] 9332.6a53

Halsey, Frederic Magie. Railway expansion in Latin America. New York. 1916. Plates. Maps. 9385.98

Harper, L. F. The coke industry of New South Wales. Sydney. 1916. Plates. Tables. [New South Wales. Geological Survey.] *9338.2944a7.23

Hodges, Le Roy. Petersburg, Virginia: economic and municipal. Petersburg. 1917. Maps. Tables. 9317.55

Indiana. Public Service Commission. Annual report. 2d. Indianapolis. 1916. Tables. *9381.09772

Ireland. Department of Agriculture and Technical Instruction. Agricultural statistics of Ireland, with detailed report for 1910–1913. London. 1911–14. 4 v. Tables. *9338.1415a7

Jevons, Herbert Stanley. The British coal trade. London. 1915. Illus. Maps. Plans. Charts. [National industries.] Bibliography, pp. 857–863. 9338.242a5

Kerr, E. The effect of wars & revolutions on government securities, external and internal. New York. 1917. Charts. Much of the book relates to the European War. 9336.3a5

King, Willford Isbell. The wealth and income of the people of the United States. New York. 1915. Charts. [The citizen's library.] 9317.3a29

Lauck, William Jett, and Edgar Sydenstricker. Conditions of labor in American industries. A summarization of the results of recent investigations. New York. 1917. Tables. Bibliography, pp. 401–403. 9331.8073a.21

Massachusetts. Bureau of Statistics. Decennial census, 1915. Bulletin. No. 1–3. Typewritten manuscript. Boston. 1917. 3 v. Tables. Two copies, one of which is kept in the Statistical Department, and one at Bates Hall Centre Desk. *9317.44a.20.1915

—Commission on the Cost of Living, 1916. Report. February, 1917. Boston. 1917. *9338.5744

National Birth-Rate Commission. The declining birth-rate: its cause and effects. New York. 1916. Charts. 9614.142a17

Norrgren, L. Russian commercial handbook. London. 1904. *9382.47a12

Ohio. Industrial Commission. Department of Investigation and Statistics. Wages and hours of labor of women and girls employed in mercantile establishments in Ohio in 1913. Columbus. 1914. Tables. Bibliography on the minimum wage, pp. 23–33. *9331.0771a6.1

Rubinow, Isaac Max. A standard accident table as a basis for compensation rates. New York. 1915. 9368.4a15

Digitized by Google

Thompson, Carl D. Municipal ownership. A brief survey of the ex rapid growth and the success of municipal ownership throughou world. New York. 1917. Tables. 9352

Todd, John Aiton. The world's cotton crops. London. 1915. P Maps. Charts. Tables. 9338.1

Willoughby, William Franklin, and others. The system of fin administration of Great Britain. [Edited by Frederick A. Cleve Introduction by A. Lawrence Lowell. Published for the Institu Government Research. New York. 1917. Tables. 9336.

Withers, Hartley. Money-changing: an introduction to foreign exch New York. 1914. Tables. 9332.

Woodbury, Robert Morse. Social insurance: an economic ans New York. 1917. Tables. [Cornell Studies in history and po science.] List of authorities cited, pp. 160–164. 9368

VITAL STATISTICS.

1917 Progress Report of the Committee on Relation of Census to Vital Statistics to the Section in Vital Statistics, American Public Health Association.

By William H. Guilfoy, M.D., and Edwin W. Kopf. *New York Medical Journal*, October 6, 1917.

This committee* was appointed in 1916 to inquire into the chief as of and requisites for public and private vital statistics work in the census year and to submit constructive recommendations on these p

This report is more of a summary of recent progress in the general of vital statistics and closely related social statistics of various kinds a program for the Census of 1920. The committee apparently thin the U. S. Census for the year 1920 in two ways, namely, as a definite for a general review and as an agency for accomplishing certain things

A goal is always desirable as a spur to effort, and there is much to b in favor of making 1920 the end of a period within which certain re should be accomplished. It should be borne in mind, however, th inventory is not only a means of closing up the accounts of past transac but is a basis for new enterprise to follow, and its taking should not be so important and absorbing as to distract attention from or to impe regular ongoing of the business.

Naturally the committee lays much stress on vital statistics, but it not neglect the opportunity to make helpful suggestions of improveme the enumeration of population, by racial and local groups, which wou of service to social investigators as well as to those working for imp health and sanitation.

* William H. Guilfoy, chairman, Cressy L. Wilbur, Frederick L. Hoffman, Wade H. Frost W. Glover, Frans Schneider, Jr., John S. Fulton, William H. Davis, Walter F. Willcox, and W. Kopf, corresponding secretary.

Digitized by Google

Perhaps the most important thing to be done at the time of or through the Census of 1920 is to obtain more nearly adequate statistics of sickness than are now available. The movement for the institution of health insurance in the various states that have already started on a program of social insurance through the enactment of workmen's compensation laws would be greatly facilitated by such statistics, and all public health work would be afforded a fact basis which it now lacks to a large degree.

The committee is to be commended for its painstaking work and it is to be hoped that some part, at least, of its excellent program may be carried out in the next census.

C. W. DOTEN.

Massachusetts Institute of Technology.

Digitized by Google

Digitized by Google

Digitized by Google

Digitized by Google

Digitized by Google